# BIOPHYSICAL STUDIES OF THE TWO N-TERMINAL COMPLEMENT CONTROL PROTEIN MODULES OF THE $GABA_B$ RECEPTOR R1a AND THEIR LIGANDS

**Stanislas C. Blein**

**A thesis submitted for the degree of Doctor of Philosophy**

**The University of Edinburgh**

**May 2002**

## ABSTRACT

The structure in solution of the N-terminal region of the $GABA_B$ receptor R1a was investigated using nuclear magnetic resonance (NMR) spectroscopy as well as other biophysical techniques. The data obtained on recombinant fragments expressed in yeast supported the presence of two complement control protein (CCP) modules. Initial findings indicated that the module pair was not amenable for structure determination.

Further work on individual modules revealed the co-existence of two stable conformers for the second CCP module that differ in the isomerisation state of Proline 119. These two isoforms are stable on the NMR time scale and do not appear to interconvert. Subsequent production of a ($^{13}C$, $^{15}N$) isotopically labelled sample of the second module allowed complete assignment of the two conformer resonances. A new protocol for structure calculation was devised to attempt simultaneous structure calculation of both conformers. The results of the structure calculation are presented.

In contrast, the first CCP module is not a compactly folded entity, despite formation of the expected disulphide bonds. It exhibits many of the characteristics of a molten globule. To ensure that this disordered state did not arise from inaccurate domain boundary prediction, different length constructs were expressed in yeast. In addition expression of this module in insect cells was attempted to rule out the possibility that the lack of ordered structure arose from a yeast-specific artefact. In parallel, the structure of the first CCP module was probed by various biophysical techniques, and the data collected showed a secondary structure content similar to that of a typical, folded CCP module. This can be considered to be consistent with a molten globule state. Interestingly, this module has some tertiary structure according to near UV circular dichroism, which could arise from disulfide-bonding.

In a yeast two hybrid system, the first module interacts with the fibulin-type module found in the fibulin extracellular matrix protein family. One possibility is that this module is functional despite its lack of compactness, and folds upon complex formation. Expression of different fragments containing the fibulin-type module from fibulin-2 was carried out using different expression systems in an attempt to assess such possibility.

## ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Paul Barlow, for his support and advice during the past four years. Especially for his help in proof-reading this thesis.

I would like to thank all of you that taught me, and answered my naïve questions. I would like to acknowledge Dr. Joanne O'Leary and Dr. Nick Mullin for teaching me some essential biochemistry. For high NMR expertise, and his help in setting-up many NMR experiments, I would like to thank Dr. Dušan Uhrín. For teaching me some basics of structure calculation throughout this work, I would like to thank Dr. Brian Smith.

On the $GABA_B$ project, I would like to thank GlaxoSmithkline for funding my final year; in particular, I thank my collaborators, Dr. Julia White and Dr. Fiona Marshall, for giving me their technical and financial support. I would also like to thank all the people that provided me assistance with the different biophysical techniques, Dr. Sharon Kelly and Dr. Alan Cooper at the University of Glasgow, and Dr. David Dryden at the University of Edinburgh.

I would like to thank the entire Barlow group for its support throughout these years. I enjoyed clubbing with Colin (a unique experience), and more seriously or not, I enjoyed any of my conversations with Andy. Thanks to Rob, Krystyna and Rosie for sharing computing skills, and for being good friends.

Finally, I sincerely want to thank Janice for her kindness and my parents for many years of support and funding.

## CHAPTER TWO: NUCLEAR MAGNETIC RESONANCE THEORY .....33

# ABBREVIATIONS

| | |
|---|---|
| AC | Adenyl cyclase |
| α-LA | α-lactalbumin |
| ANS | 8-anilino-1-naphtalenesulfonate |
| APP | β amyloid precursor protein |
| BMD | Buffer minimal dextrose |
| βME | β-mercaptoethanol |
| BMG | Buffer minimal glycerol |
| BMM | Buffer minimal methanol |
| bp | Base pair |
| cAMP | Cyclic AMP |
| CD | Circular dichroism |
| cDNA | Complementary DNA |
| CNS | Central nervous system |
| COSY | Correlated spectroscopy |
| CP, CP-module, CCP | Complement control protein module |
| CS R | Ca2+ sensing receptor |
| DPFGSE | Double pulsed field gradient spin echo |
| DSC | Differential scanning calorimetry |
| DTT | Dithiothreitol |
| ECM | Extracellular matrix |
| EDTA | Ethylenediamine tetra-acetic acid |
| EndoH | Endoglycosidase H |
| $EndoH_f$ | Endoglycosidase $H_f$ |

| | |
|---|---|
| FFT | Fast fourier transformation |
| FID | Free induction decay |
| FPLC | Fast protein liquid chromatography |
| FT | Fourier transformation |
| GABA | Gamma-aminobutyric acid |
| $GABA_B$ R1 | Metabotropic GABA receptor subunit 1 |
| $GABA_B$ R2 | Metabotropic GABA receptor subunit 2 |
| GdnHCl | Guanidine hydrochloride |
| GlcNAc | N-acetylglucosamine |
| GSH | Reduced glutathione |
| GSSG | Oxidized glutathione |
| GST | Glutathione S-transferase |
| HIC | Hydrophobic interaction chromatography |
| His6 | Hexa-histidine |
| HMQC | Heteronuclear multi quantum spectroscopy |
| HOHAHA | Homonuclear Hartmann-Hahn spectroscopy |
| HPLC | High-pressure liquid chromatography |
| HSQC | Heteronuclear single quantum spectroscopy |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| LB | Luria-bertani |
| LCMS | Liquid chromatography mass spectrometry |
| LIVBP | Leucine/isoleucine/valine-binding protein |
| LTP | Long term potentiation |
| MD | Minimal dextrose |
| MEM | Maximum entropy method |

| | |
|---|---|
| mGlu R | Metabotropic glutamate receptor(s) |
| mGlu R1 | Metabotropic glutamate receptor type 1 |
| NMR | Nuclear magnetic resonance |
| NOE | Nuclear Overhauser effect |
| NOESY | Nuclear Overhauser effect spectroscopy |
| NTER | N-terminal domain |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| pdb | Protein databank |
| PFG | Pulsed magnetic field gradients |
| PMSF | Phenylmethanesulfonyl fluoride |
| PVDF | Polyvinyldifluoride |
| Q-TOF | Quadrupole time of flight |
| RMSD | Root mean square deviation |
| ROE | Rotating Overhauser Effect |
| RP | Reverse-phase |
| SCUBA | Stimulated cross-relaxation under bleached alphas |
| SDS-PAGE | Sodium dodecylsulphate polyacrylamide gel electrophoresis |
| SE | Sensitivity enhanced |
| TCA | Trichloroacetic acid |
| Tdx | Thioredoxin |
| TFA | Trifluoroacetic acid |
| TMD | Trans-membrane domain |
| TOCSY | Total correlation spectroscopy |
| TPPI | Time proportional phase incrementation |

| | |
|---|---|
| VCP | Vaccina virus complement control protein |
| YPD | Yeast extract peptone dextrose |
| YPDS | Yeast extract peptone dextrose sorbitol |
| YTH | Yeast two hybrid |

## ABBREVIATIONS OF THE DIFFERENT RECOMBINANT PROTEIN FRAGMENTS

| | |
|---|---|
| CP1 | $GABA_B$ R1a complement control protein module one |
| WT-CP1 | $GABA_B$ R1a complement control protein module one<br>Residues 17-98 |
| TR-CP1 | $GABA_B$ R1a complement control protein module one<br>Residues 24-98 |
| CP2 | $GABA_B$ R1a complement control protein module two<br>Residues 96-159 |
| CP12 | $GABA_B$ R1a complement control protein module pair |
| WT-CP12 | $GABA_B$ R1a complement control protein module pair<br>Residues 17-159 |
| FLM-CP12 | $GABA_B$ R1a complement control protein module pair, mutated for the first N-glycosylation site (Asn23 mutated to Ala)<br>Residues 17-159 |
| TR-CP12 | $GABA_B$ R1a complement control protein module pair<br>Residues 24-159 |
| EGFs~FTM2 | Epidermal growth factor-like repeat 8-10 and fibulin-type module from fibulin2<br>Residues 942-1184 |
| EGF~8-10 | Epidermal growth factor-like repeat 8-10 from fibulin2<br>Residues 942-1072 |
| FTM2 | fibulin-type module from fibulin2<br>Residues 1069-1184 |

## NAMES, ABBREVIATIONS AND SINGLE LETTER CODES FOR AMINO ACIDS

| | | |
|---|---|---|
| Alanine | Ala | A |
| Arginine | Arg | R |
| Asparagine | Asn | N |
| Aspartic acid | Asp | D |
| Cysteine | Cys | C |
| Glutamic acid | Glu | E |
| Glutamine | Gln | Q |
| Glycine | Gly | G |
| Histidine | His | H |
| Isoleucine | Ile | I |
| Leucine | Leu | L |
| Lysine | Lys | K |
| Methionine | Met | M |
| Phenylalanine | Phe | F |
| Proline | Pro | P |
| Serine | Ser | S |
| Threonine | Thr | T |
| Tryptophan | Trp | W |
| Tyrosine | Tyr | Y |
| Valine | Val | V |

# CHAPTER ONE: INTRODUCTION

## 1.1 The metabotropic $GABA_B$ receptor

Gamma-aminobutyric acid (GABA) is the principal inhibitory neurotransmitter in mammalian brain. There are two distinct categories of target receptors for GABA each of which mediates synaptic inhibition: the ionotropic GABA-types A and C ($GABA_{A/C}$) receptors; and the metabotropic GABA-type B ($GABA_B$) receptors.

Upon activation by two molecules of GABA, the pentameric $GABA_A$ receptor induces a rapid inhibition of neuronal electrical activity. This comes about because the $GABA_A$ receptor is an ion channel which opens upon appropriate ligation allowing the inward flux of $Cl^-$ and a subsequent hyperpolarisation of the postsynaptic membrane [1]. In contrast, activation of postsynaptic $GABA_B$ receptors produces a slower, more prolonged inhibition [2]. Presynaptic $GABA_B$ receptors also exist and these may act as autoreceptors by suppressing GABA release from GABA-ergic nerve terminals, or they may act as heteroreceptors by blocking the release of other neurotransmitters, including L-glutamate (the main excitatory neurotransmitter), from the appropriate nerve terminals [3]. These effects are believed to be mediated by coupling of $GABA_B$ receptors to the heterotrimeric guanine-nucleotide binding (G)-proteins, Gi and Go. $GABA_B$ receptors are thus able to modulate adenylyl cyclase activity, and to cause inwardly rectifying $K^+$ channels to open and voltage-dependent $Ca^{2+}$ channels to close [4, 5] (Figure 1.1).

$GABA_B$ autoreceptors in the hippocampus can regulate the induction of the long-term 'potentiation' (LTP) of synaptic transmission [6]. The phenomenon of LTP represents an important model system for studies of associative enhancement of synaptic strength as it is thought to underlie cognitive processes such as learning and memory. Other physiological actions have also been attributed to the stimulation of $GABA_B$ receptors, including analgesia, modulation of respiratory activity, and muscle relaxation [2]. $GABA_B$ receptors are thought to be potential therapeutic targets in a range of conditions. Lioresal, a racemic mixture of baclofen, is a receptor agonist used clinically in the treatment of spasticity arising from multiple sclerosis or spinal injury [7]. Other possible clinical applications of $GABA_B$ agonists/antagonists include epilepsy [2, 8], drug withdrawal [9, 10]. anxiety, depression, hypertension, and cognitive dysfunction [4, 11].

The $GABA_B$ receptor was first classified and pharmacologically characterised by Hill and Bowery in 1981 [12]. Much additional pharmacological and biochemical information was gleaned over the subsequent decade-and-a-half. Despite this, the cloning of a gene encoding a $GABA_B$ receptor proved very difficult. The much awaited breakthrough came in 1997 with the first report of the successful cloning of cDNA encoding a $GABA_B$ receptor subunit ($GABA_B$ receptor type 1, $GABA_B$ R1). This work, carried out by a team led by Bernhard Bettler at Novartis in Basel [13], represents an important landmark. Significant progress in our understanding of the structure and function of the $GABA_B$ receptor has ensued rapidly, revealing several surprising and possibly unique aspects of the receptor. In the original report of the cloning work, two splice variants, named $GABA_B$ R1a and $GABA_B$ R1b, were identified. Subsequently, the gene encoding a second subunit of the receptor ($GABA_B$ R2) was cloned [14-20]. Both the R1a/b and R2 subunits are members of the family of G-protein coupled receptors possessing seven transmembrane domains. Only when $GABA_B$ R1 (either R1a or R1b) and $GABA_B$ R2 are co-expressed is full activity obtained, and the $GABA_B$ receptor is a heterodimer composed of $GABA_B$ R1a/b and $GABA_B$ R2 subunits (Figure 1.1).

Despite these successes, the number of receptor genes cloned to date may not be sufficient to explain the results of numerous studies obtained over the last 15 years which have suggested considerable heterogeneity of $GABA_B$ receptors [2, 21]. It is not unlikely that other splice variants, and perhaps even additional subunits remain unidentified. The area is one of intense activity.

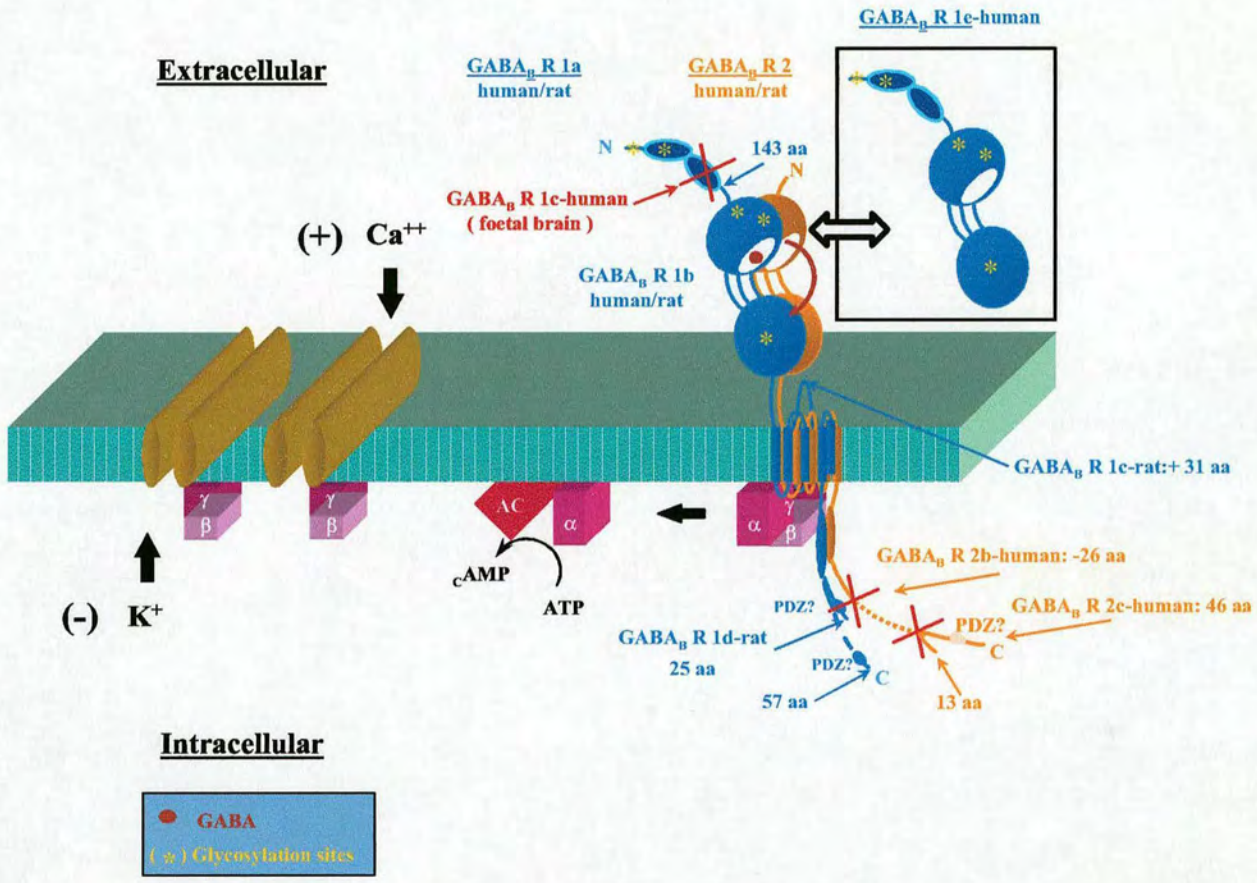



**Figure 1.1: The GABA$_B$ heterodimer and splice variants**

**Figure 1.1 (continued):The $GABA_B$ heterodimer and splice variants**

The $GABA_B$ R1 subunit is a protein of approximately 100 kDa. In the $GABA_B$ R1b subtype (human/rat), the N-terminal extracellular domain is the ligand binding domain and consists of a region with sequence homology to a family of bacterial periplasmic binding proteins (LIVBP).

The $GABA_B$ R1a splice variant (human/rat) differs at the N-terminus by the presence of a tandem pair of CP modules, the $GABA_B$ R1c-human splice variant differs from $GABA_B$ R1a in that is missing the second CP module, while the $GABA_B$ R1c-rat differs from R1 in the 5th transmembrane region and the 2nd extracellular loop by the insertion of an additional 31 amino-acids.

$GABA_B$ R1d-rat differs in the intracellular C-terminal domain where an insertion of 566 base pairs that includes a stop codon could generate a truncated receptor.

The $GABA_B$ R1e-human subtype represents a soluble isoform of the $GABA_B$ R1a splice variant, which doesn't seem to bind GABA or its analogues.

Note that the $GABA_B$ R1a/b subunit has a consensus motif for PDZ domain binding at its C-terminus.

The $GABA_B$ R2 subunit (human/rat) is 35% identical at the amino acid level to the $GABA_B$ R1 subunit but has a longer intracellular C-terminal tail (80 extra residues) compare to $GABA_B$ R1a/b, and has no affinity for GABA or its analogues

$GABA_B$ R2 splice variants include $GABA_B$ R2b-human which contains a deletion of 26 amino acids at the C-terminus whereas $GABA_B$ R2c-human diverges 13 amino acids also from the C-terminus and as a result has a different C-terminus of 46 amino acids.

Contact between the two subunit C-terminus seems to be mediated by a "coiled-coil" region in each subunit and is needed for receptor trafficking. $GABA_B$ R1 "coiled-coil" interacts directly with the transcription factors CREB2 and ATFx.

The dimerization motif is contained within the extracellular binding domain. Upon ligand binding it has been suggested that the amino-terminal domain undergoes a hinge-bending motion, as in LIVBP described as a "venus fly trap" mechanism.

Receptor activation causes adenylyl cyclase inhibition. This effect is believed to be mediated by the $\alpha$ subunit of Gi/o type G proteins. Postsynaptic receptors cause inwardly rectifying $K^+$ channels to open allowing $K^+$ to move down its electrochemical gradient, while presynaptic receptors cause voltage-dependent $Ca^{2+}$ channels to close. The effect on $K^+$ channels is mediated by the $\beta/\gamma$ subunits of Gi/o type G proteins and effect on $Ca^{2+}$ channels is mediated by the $\beta/\gamma$ subunits of Go type G protein [22].

### 1.1.1 Cloning of the first subunit and first discovered splice variants

The key to the first successful cloning of complementary DNA for a $GABA_B$ receptor was the development of the high affinity antagonist [125I]CGP64213 [13]. Such antagonist was developed by Novartis, and has activity at pre- as well as postsynaptic $GABA_B$ receptors, as shown by complete suppression of L-baclofen-induced responses in electrophysiological recordings from rat CA1 hippocampal slices [13]. Subsequent photoaffinity labelling revealed two candidate receptor glycoproteins with molecular weight of 100 kDa and 130 kDa (90 kDa and 110 kDa respectively after treatment with N-glycosidase) in the cortex, cerebellum and spinal cord of human, rat, mouse, chicken, frog and zebra fish. Expression cloning was then used to identify a 4.4 kb cDNA insert in a rat cortex and cerebellum library that had been transfected into COS-1 cells. The first gene to be identified encodes a 960 amino acid protein ($GABA_B$ R1a) including a 16-residue signal sequence and corresponds, after post-translational N-glycosylation, to the 130 kDa glycoprotein photoaffinity labelled in native brain tissue. Subsequent low-stringency hybridisation cloning using the $GABA_B$ R1a cDNA as a probe identified a 2.9 kb cDNA which encodes a protein corresponding to the 100 kDa glycoprotein ($GABA_B$ R1b), and which represents a splice variant of $GABA_B$ R1a. mRNA for

both variants is abundant in brain, and transcripts for the receptor are found in all cerebral cortical areas, the pyramidal cell layers of the hippocampus, the granular cell layers of the dentate gyrus, and basal ganglia. As confirmation that $GABA_B$ R1a/b indeed correspond to variants of a functional $GABA_B$ receptor, the cloned proteins, heterologously expressed in Human Embryonic Kidney (HEK) 293 cells, were shown to couple negatively to adenylyl cyclase. Although when expressed heterologously in cell lines, the antagonist pharmacology of the cloned receptor was shown to be similar to that of native receptors, the potency of agonist ligands was about 100-fold less than expected from studies of brain membranes [13].

Human homologues of the two proteins were later reported, and share 99% sequence identity with the rat proteins [23].

### 1.1.2 Classification of $GABA_B$ R1a and 1b, and sequence similarities

The primary sequence of $GABA_B$ R1, and analysis of its hydropathy plots, places it within the superfamily of G-protein coupled receptors, all of which encompass seven putative transmembrane domains (7TMD) [13]. There are five classes of 7TMD receptors [24], each containing proteins that are related by sequence similarity, but there is no significant similarity between classes. Class A members are most numerous and related to rhodopsin; class B members are related to the vasoactive intestinal peptide and glucagon receptors; class C receptors are homologous to the metabotropic glutamate receptors (mGlu R) and include the $Ca^{2+}$ sensing receptor (CS R); class D contains the pheromone receptors; and class E contains the cAMP receptors of *Dictyostelium* [24]. The sequence of $GABA_B$ R1a/b is 18-23% identical to the eight subtypes of mGlu R and therefore it is grouped in class C [13]. The mGlu R subtypes are receptors for the major excitatory neurotransmitter [25], while CS R is involved in calcium homeostasis in humans and other mammals [26]. The similarity with mGlu R extends over the length of the sequence, and is striking since 27 out of 208 residues conserved within all mGlu R subtypes are present in $GABA_B$ R1. Seventeen of these are in the ~550 residue N-terminal, extracellular domain of $GABA_B$ R1. $GABA_B$ R1, however, lacks a domain present in the mGlu R's and in the CS R that contains nine closely spaced cysteines [27].

All class C G-protein coupled receptors share a similar molecular architecture consisting of a large extracellular amino-terminal domain encompassing a ligand-binding site [28-30], followed by seven transmembrane domains and the intracellular C-terminal tail. Coupling to G-proteins is mediated by

the intracellular loops connecting the seven transmembrane domains, and the C-terminal region. The short, third intracellular loop, which is highly conserved among class C 7TMD receptors and is essential for G-protein activation [31].

There is a marked similarity between the N-terminal domains of class C members and a family of bacterial periplasmic amino acid-binding proteins which include the leucine-binding protein (LBP) and the leucine/isoleucine/valine-binding protein (LIVBP) [32] (Figure 1.2 & 1.3). These binding proteins are found in Gram-negative bacteria and are essential to the uptake of low molecular weight compounds such as amino-acids and sugars. The recognition and specific binding of these molecules is an essential first step for transmembrane transport. This theme is elaborated further below since the similarity has been used as a basis for construction of a structural and functional model of the N-terminal domain of $GABA_B$ R1a [33].

There is also a striking similarity between the N-terminal 143 residues of $GABA_B$ R1a and proteins of the mammalian complement system (Figure 1.1 & 1.2) [34]. The complement protein (CP) category of protein modules (also called sushi domains) was first identified [and called the short consensus repeat (SCR)] [35] as a common feature in the sequences of the family of proteins that regulate complement activation. In $GABA_B$ R1b, these 143 residues are replaced by an unrelated sequence of 18 amino acids.

Finally, the C-terminus of $GABA_B$ R1a/b contains several features of interest. There are consensus substrate sites for casein kinase and protein kinase C suggestive of modulation by phosphorylation [13]. There is also a "coiled-coil" sequence of 35 residues (Figure 1.1) [16, 36] and a putative PDZ recognition sequence (SRV) at position –6 from the C-terminus. (Figure 1.1) [5]. As seen for NMDA and AMPA receptors, which C-termini are involved in synaptic targeting through interactions with PDZ-domain containing adaptor proteins. These proteins, in turn, have additional binding sites that allow interactions with the cytoskeleton and with signal transduction molecules [37].

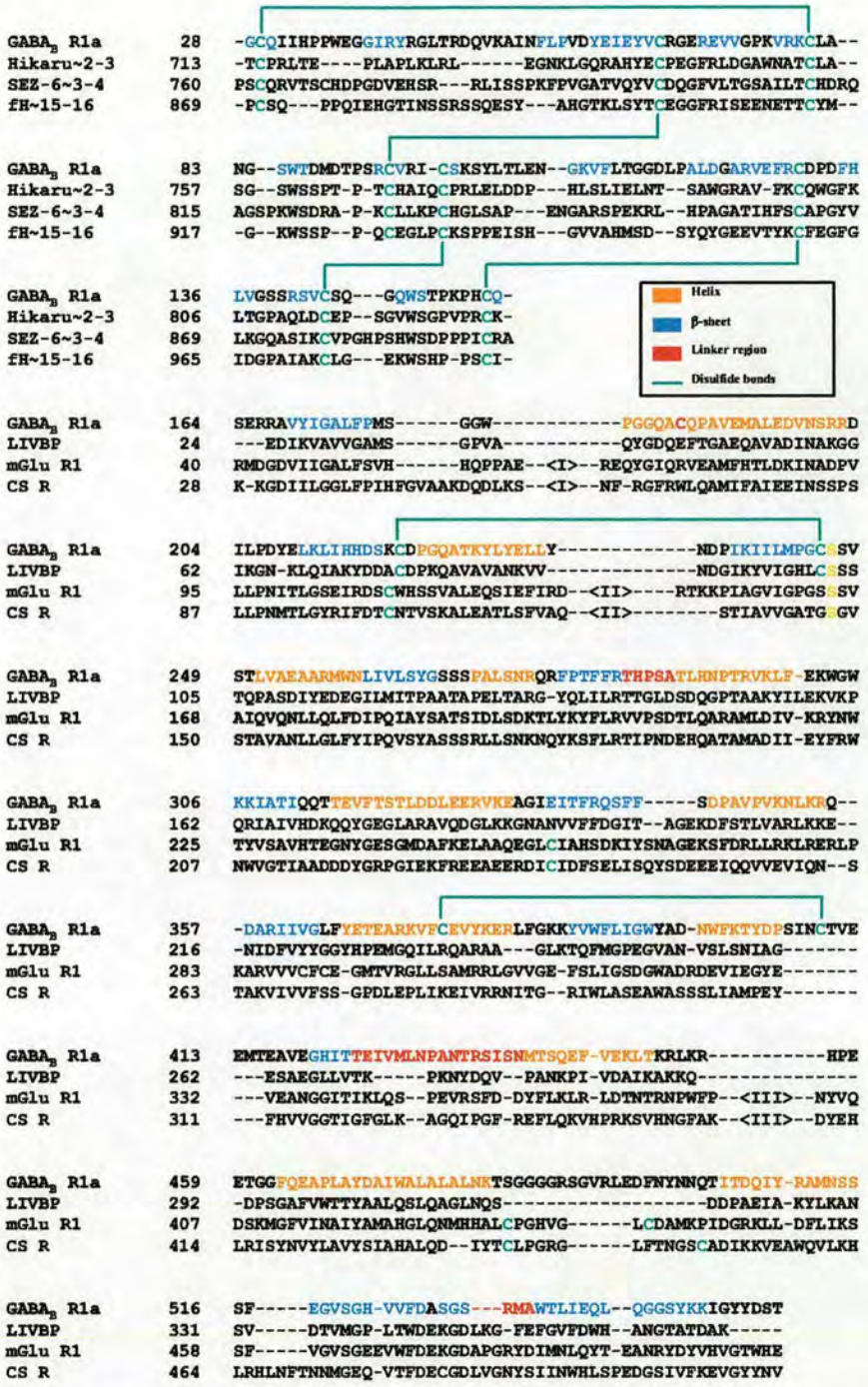



**Figure 1.2:** **Top** : An alignment of the $GABA_B$ R1a pair of CP-modules with CP-pairs found in the CNS (the 2nd and 3rd CPs from the *Drosophila melanogaster* Hikaru genki protein, and the 3rd and the 4th CPs from the rat Sez-6 protein), and a well characterised CP-pair found in the complement system (the 15th and the 16th CPs from the human factor H protein).

**Figure 1.2 (continued):** **Bottom**: alignment of the extracellular domain from the rat $GABA_B$ R1a with the LIVBP (*E. coli*), the extracellular domain from the rat mGlu R1, and the extracellular domain from the rat CS R according to Galvez et al. (1999) [33] and Brauner-Osborne et al. (1999) [30]. Note that Cys101 and Cys 236 of CS R align with Cys109 and Cys254 of the mGlu R1 respectively - two residues involved in the covalent, disulphide-linked homodimerization of the CS R [38].

Amino acid numbering includes signal peptides, three insertions found in the N-terminal domain of mGlu R1 have been removed and indicated by <I>, <II>, and <III> [33].

**Hyphens in the sequences indicate gaps introduced for optimal alignment.**

### 1.1.3 Cloning of a second subunit - the metabotropic GABA receptor is a heterodimer

Despite the convincing demonstration by Kaupmann *et al.* (1997) [13] that $GABA_B$ R1a/1b fulfilled many of the criteria of a functional metabotropic GABA receptor there remained some discrepancies between properties observed for the recombinant proteins on the one hand, and the native brain receptor on the other [13, 23]. Agonists showed significantly lower potencies for the recombinant receptors. This was attributed, at least in part, to inefficient coupling to G-proteins within the heterologous expression system tested. Indeed, coupling to adenylyl cyclase was weaker than anticipated and coupling to ion-channels difficult to demonstrate. Subsequently, Couve *et al* [39] showed that heterologously expressed recombinant $GABA_B$ R1a/b were not transported efficiently to the cell surface and that these proteins were instead largely retained in the endoplasmic reticulum. There was thus considerable excitement when five groups simultaneously reported the existence of a second subtype of the $GABA_B$ receptor ($GABA_B$ R2) [14-18], (later reported by a sixth group as GPR51 [19, 20]). Not only did this discovery resolve many of the earlier difficulties raised by the work of Kaupmann *et al.*[13, 23], and Couve *et al.* [39], but it also opened a new chapter in the story of 7TMD receptors.

The discovery of a second subunit was made in 1998 using a combined bioinformatics and molecular biology approach. The human homologues of $GABA_B$ R1a/b had been identified and shown to be 98% identical to the rat sequences. The human cDNA sequence was then used to mine a database of expressed sequence tags ultimately revealing a second receptor subunit ($GABA_B$ R2). The $GABA_B$ R2 subunit has 941 residues and is 35% identical and 54% similar to $GABA_B$ R1b. The N-terminal extension present in $GABA_B$ R1a is not found in the R2 subunit. Furthermore, the R2 subunit has an unusually long, intracellular C-terminal tail containing 80 residues more than the equivalent region of $GABA_B$ R1a/b (Figure 1.1).

$GABA_B$ R2 assembles with $GABA_B$ R1 by means of parallel coiled-coil a-helices located in the respective carboxy-terminal intracellular domain of each subunit as demonstrated by yeast-two-hybrid (YTH) technique. Further studies showed that the intracellular carboxy-termini of $GABA_B$ R1 contain an endoplasmic reticulum retention signal that prevents the immature, non-glycosylated protein from being expressed at the cell surface [40-42]. The coiled-coil interaction with the corresponding carboxy-terminus of $GABA_B$ R2 masks the signal and enables the release of $GABA_B$ R1 from the ER [42]. $GABA_B$ R2 which itself is expressed at the cell surface, therefore serves to chaperone $GABA_B$ R1 to the cell surface.

However, $GABA_B$ R1 that lacks the ER retention signal is not sufficient to form a fully functional receptor, and heterodimerisation is thus required to form a fully functional $GABA_B$ receptor, the $GABA_B$ receptor is an heterodimer made of two subunits: $GABA_B$ R1 and $GABA_B$ R2.

Unlike $GABA_B$ R1, $GABA_B$ R2 does not bind either agonists or antagonist with measurable potency but when the two subunits were co-expressed in a 1:1 ratio in HEK293T cells [14], binding studies revealed agonist potencies comparable with the native brain receptors and coupling to G-proteins. A ten-fold increase in binding potency is observed, as measured by the inhibition of antagonist [$^{125}$I]CGP64213 binding to receptor by agonists [15]. Co-expression of $GABA_B$ R1, $GABA_B$ R2 (in combination with $G_{\alpha 01}$, a G-protein $\alpha$-subunit) in HEK293T cells resulted in agonist-dependent stimulation of [35S]GTP-$\gamma$S binding [16]. Finally, and critically, when co-expressed in *Xenopus* oocytes the two subunits produced a GABA-sensitive coupling with inwardly rectifying $K^+$ channels [14-17].

Early reports suggested that recombinant $GABA_B$ R2 could display some functionality [17, 18]. Although it now appears unlikely that wild type $GABA_B$ R2 subunit can function alone. The lack of detectable $GABA_B$ response in tissue from mice lacking the gene for $GABA_B$ R1 subunit argues

against $GABA_B$ R2 as an autonomous receptor. $GABA_B$ R2 is also heavily down-regulated in such null-mutant mice, a finding which support the notion that in wild-type mice, most of the $GABA_B$ R2 protein is associated with $GABA_B$ R1, in agreement with biochemical studies [43]. The same biochemical studies strongly suggest that $GABA_B$ R1a and $GABA_B$ R1b do not homo- or hetero-dimerize and that $GABA_B$ R1a, $GABA_B$ R1b, and $GABA_B$ R2 exist exclusively in heteromeric form in the rat brain.

Further evidence for the formation of a heterodimer has been provided by peptides derived form the carboxy-terminal domains of $GABA_B$ R1 and $GABA_B$ R2 that form heterotypic parallel α-helical coiled-coil domain structures when combined in vitro [36]. Upon sedimentation equilibrium the $GABA_B$ R1/ $GABA_B$ R2 coiled-coil domain peptides assemble in a 1:1 stoichiometry.

$GABA_B$ R1 null mice exhibit spontaneous seizures, hyperalgesia, hyperlocomotor activity, and memory impairment [44]. Such findings suggest that use of $GABA_B$ agonists may be of benefit in the treatment of psychiatric and neurological disorders in which attentional processing is impaired, and epilepsy.

### 1.1.4 Splice variants amongst the subunits of the metabotropic GABA receptor

More splice variants of the $GABA_B$ receptor subunits have now been described, both in human and rat. However most variants have not been fully characterised in both terms of function and localisation, some are unpublished and have simply been reported through Genbank (Human $GABA_B$ R1c, Human $GABA_B$ R2b, Human $GABA_B$ R2c).

Two splice variants, $GABA_B$ R1c and $GABA_B$ R1d were subsequently identified in rodent [45]. $GABA_B$ R1c is identical to $GABA_B$ R1b but with an insertion of 93 base pairs (bp) that generates an additional 31-amino-acids in the second extracellular loop and fifth transmembrane region, between Gly~654 and Ile~655 (Figure 1.1). Note that the fifth transmembrane domain is, in general, the most solvent-exposed domain according to modelling studies of 7TMD proteins [46] and is conceivably available for protein and drug interaction. For example, a non-covalent dimerization consensus motif ( **L**MAL**G**FLI**G**YTC**L**, consensus residues in bold), originally found in the unrelated ß-adrenegic receptors, is present in the fifth transmembrane domain of the CS R, and has been proposed to mediate homodimerisation via hydrophobic interactions [47]. When co-expressed with $GABA_B$ R2, $GABA_B$ R1c is targeted to the cell membrane, and gives rise to a functional heterodimer

with similar properties to those of the $GABA_B$ R1a - $GABA_B$ R2 heterodimer [48]. The $GABA_B$ R1d variant has an additional insertion of 566 bp that generates a divergent C-terminal region (Figure 1.1). This insertion includes a stop codon that results in the replacement of 57 residues of R1b with 25 residues of unrelated sequence. No functional studies on such variant have been reported to date.

The human $GABA_B$ R1c-splice variant identified by Marshall and colleagues (unpublished – Genbank accession number AJ012187) is unrelated to rat $GABA_B$ R1c and differs from $GABA_B$ R1a in that it has a 62 amino acids deletion that removes the second CP module. This subtype function was never been investigated; only RT-PCR studies have shown a up-regulation of this variant in the foetal brain [49].

The $GABA_B$ R1e subtype represents a soluble isoform of the $GABA_B$ R1a variant detected in both rats and humans, encoding only the CP-module pair and the extracellular ligand binding domain [50]. When expressed in heterologous system, this truncated receptor is both secreted and membrane associated but lacks the ability to bind radiolabeled antagonist ([$^3$H] CGP 54626A), activate GIRK channels, and inhibit forskolin-induced cAMP production, either alone or co-expressed with $GABA_B$ R2. Interestingly when co-expressed with $GABA_B$ R2 the truncated receptor is able to disrupt the $GABA_B$ R1a/$GABA_B$ R2 association but doesn't disrupt GIRK channels activation by the full-length heterodimer. Because $GABA_B$ R1e lacks an intracellular domain and therefore would not be able to dimerize via a coiled-coil interaction, it was speculated that $GABA_B$ R1e and consequently $GABA_B$ receptors could dimerize using a motif located within the amino-terminal domain like other metabotropic receptors (mGlu R). To determine the region of $GABA_B$ R1e responsible for dimerization, Schwarz et al used a series of progressive N-terminal truncations to show that the critical domain resides at the beginning of the ligand-binding domain (residues 166-439, the numbering system used here includes the 16 residues of signal sequence) and is outside the limits of the CP-module pair.

Human $GABA_B$ R2 splice variants include two human splice forms, $GABA_B$ R2b (unpublished –Genbank accession number AF095723) contains a deletion of 26 amino acids at the C-terminus whereas $GABA_B$ R2c (unpublished –Genbank accession number AF095724) diverges 13 amino acids from the C-terminus and has a different C-terminal tail of 46 amino acids [49]. The functional properties of these variants are currently unknown.

### 1.1.5 Expression profiles of the $GABA_B$ receptor subunits and splice variants

#### 1.1.5.1 Postsynaptic versus presynaptic

The initial identification of the $GABA_B$ R1a and R1b splice variants led to the rather attractive hypothesis that the R1a and R1b variants might provide the basis for the distinction between presynaptic, and postsynaptic, $GABA_B$ receptors [5]. As these two splice variants have identical intracellular regions, a differential subsynaptic association would presumably be due to differential membrane targeting or co-assembly with yet unidentified membrane proteins capable of distinguishing between the extracellular domains of R1a and R1b.

A focal point for the investigation of this issue has been the rat cerebellum where $GABA_B$ binding sites are at their highest density in the classically termed "molecular layer" of the cerebellum. Granule cell bodies located in the morphologically distinguishable granule cell layer send parallel axon fibres through this "molecular layer" where they form excitatory presynaptic nerve terminals impinging upon the dendrites of Purkinje cells. The Purkinje cell bodies themselves form a highly characteristic morphological feature, the Purkinje cell layer. *In situ* hybridization studies were performed to determine whether there was a differential distribution of $GABA_B$ R1a and R1b mRNA in this region of the brain [51]. $GABA_B$ R1a mRNA was found to be expressed predominantly in the granule cell layer (approximately 16-fold higher grain density than in the Purkinje cell layer) whereas R1b mRNA was localized primarily to Purkinje cell bodies (approximately 8-fold higher grain density than in the granule cell layer). These results predict that R1a-containing receptor proteins are preferentially expressed in granule cells, the source of the presynaptic terminals, while on the other hand, $GABA_B$ R1b-containing receptor proteins are differentially produced in the postsynaptic Purkinje cells, whose dendrites form synapses, albeit excitatory ones, with the granule cell terminals. These observations appear to support the hypothesis of a differential association of $GABA_B$ R1a with presynaptic structures and R1b with postsynaptic membranes.

In a biochemical approach to the question of differential presynaptic versus postsynaptic association of $GABA_B$ R1a and R1b splice variants [43], a subcellular fraction of postsynaptic densities (PSD) from rat brain membranes was prepared and enriched by differential centrifugation. Immunoreactivity against both R1a and R1b was observed in fractions containing synaptic plasma membranes, suggesting a significant synaptic localisation for both splice variants *in vivo*. A further purified PSD fraction, however, which is enriched in postsynaptic membranes and associated

submembranous proteins appear deficient in R1b immunoreactivity but give a strong signal with the antibody specific for R1a. These results suggest, contrary to earlier expectations, that it is the R1a splice variant which may be preferentially associated with postsynaptic structures. In summary, the question of differential localisation of R1b and R1a splice variants remains unresolved, although biochemical studies are favoured as mRNA quantification is not always reliable.

### 1.1.5.2 Localisation in the CNS and peripheral tissues

$GABA_B$ R1 splice variants are expressed ubiquitously in rat and human either in the CNS or peripheral tissues. The most interesting differences are observed in the uterus, spleen and lung, in which the sum of the R1 subunit expression is higher than in many areas of the CNS [49]. $GABA_B$ R2 splice variants are also expressed throughout the CNS, with the possible exception of the pituitary gland, and at relatively lower levels in the basal ganglia (striatum), white matter tracts (corpus callosum) and spinal cord but strikingly are almost completely absent in all human peripheral tissues.

$GABA_B$ R1 and R2 are also detected in cardiomyocytes where they can be activated and induce electrophysiological alterations through Kir channels [52]. In man, hypertension is associated with the activation of the sympathetic nervous system but the mechanisms for this are unclear. $GABA_B$ receptors have been implicated at several levels. GABA-ergic axons are found throughout superior cervical (sympathetic) ganglia consistent with a neuromodulatory role for GABA in the peripheral nervous system [53]. Also there is evidence that spinal $GABA_B$ receptors regulate sympathetic vasomotor tone [54]. Likewise, alterations in central nervous system $GABA_B$ receptors have been implicated in the regulation of cardiovascular function in hypertension.

### 1.1.6 Structural insights into the metabotropic $GABA_B$ receptor

On the basis of sequence comparisons, the expression and characterisation of truncated fragments of $GABA_B$ R1a/b, and construction of chimeric receptors, a picture of the global architecture of $GABA_B$ R1 is emerging (Figure 1.1). To date, there are no differences in the pharmacological profile of $GABA_B$ heterodimers containing alternatively spliced forms of $GABA_B$ R1 subunit [13, 28, 33, 48].

#### 1.1.6.1 CP-modules of $GABA_B$ R1a

The 143 residues distinguishing $GABA_B$ R1a from $GABA_B$ R1b form a tandemly arranged pair of CP-modules [34]. This is the first example of CP-modules occurring in a seven transmembrane domain receptor. There are two other 7TM receptors known to possess extracellular modules. These belong to class B, the secretin receptor family, and have various numbers of EGF-like modules, rather than CP-modules, in their large N-terminal domains [55]. In the case of one of these receptors, the leukocyte activation antigen, CD97, alternative splicing results in three, four or five EGF-like modules and a corresponding modulation in affinity for its ligand [56].

The CP-modules are the predominant module type within several cell surface complement regulatory proteins with single transmembrane domains such as CD46 (which is the measles virus receptor) and CD35. They are also found in a GPI-linked complement regulator (CD55) and in soluble complement proteins. In addition CP-modules have been found in many non-complement proteins including the IL2β receptor. There are several other examples of CNS proteins that contain CP-modules (Figure 1.2) including the *Drosophila* protein "hikaru genki" which has four CP-modules [57], the human protein, neurocan, which has a single CP-module [58], and mouse SEZ-6 a single transmembrane domain protein possessing five CP-modules [59, 60]. SEZ-6 was identified as a protein whose expression was enhanced by the perfusion of brain slices with convulsant drugs; the human equivalent is a hypothetical 87.6 kDa foetal brain protein which is 49% identical to the C-terminal 746 residues of the SEZ-6 protein. A CP-module containing protein has also been found in the retina [61]. Although some examples of CP-modules function as structural or spacer units in bigger proteins, wherever they occur towards the N-terminus of a well-studied cell-surface protein, they have been shown to participate in specific protein-protein interactions. It was therefore speculated that the N-terminal CP-modules of $GABA_B$ R1a might prove to be of importance in

recognising another, extracellular or membrane-bound protein. This putative interaction might be important for anchoring the receptor at a particular cellular location, or might have a modulatory influence on receptor activity and signal transduction. The only 7TM receptor known to possess a cellular ligand is CD97, mentioned earlier. As it happens, its ligand has been identified as CD55 which is a GPI-anchored CP-module containing protein [62]. The interaction is between an EGF-like module of CD97 and the N-terminal CP-module of CD55 [63].

A fragment of $GABA_B$ R1a corresponding to residues 24-159 has been recombinantly expressed in *Pichia pastoris*, purified to homogeneity and shown to adopt a stable, soluble fold in isolation from the rest of the protein [34]. Sequence comparisons (Figure 1.2), differential scanning calorimetry, and circular dichroism were consistent with the presence of two CP modules (also called short consensus repeats, CCPs or sushi domains) separated by a linker sequence of three amino-acids - CP module 1 extends from residue 28-95, and CP module 2 from residue 99-158 [34].

On the basis of homology with a tandem pair of CP-modules for which a structure had been solved by NMR, a model was built for residues 28-158 [34]. This indicated that each module adopts a typical CP-module like fold in which a compact hydrophobic core containing highly conserved residues is sandwiched between antiparallel β-sheets made up from short β-strands. Four cysteines, disulphide linked in the pattern 1-3 (i.e. Cys~29-Cys~80/Cys~99-Cys~144) and 2-4 (i.e. Cys~66-Cys~95/Cys~130-Cys~156), stabilise each module. Module 1 is a less typical example of a CP-module than module 2 and has a large insertion of 12 residues (Arg~43-Asn~54) that appears as a looped out segment in the model. This is also an N-terminal extension of seven residues in module 1 that are not part of the CP-module consensus sequence and include an N-glycosylation site. On the basis of precedent, the two modules are probably joined flexibly in an end-to-end fashion generating a highly extended structure with a large surface area available for binding with another domain within $GABA_B$ R1a, or with another protein. N-glycosylation sites are surface-exposed at positions Asn~23 and Asn~83 and therefore exclusively located on module 1 (numbering from rat $GABA_B$ R1a includes signal sequence).

### 1.1.6.2 GABA binding domain

The sequence immediately beyond the distinct N-terminal sequences of $GABA_B$ R1a/b, and prior to the first putative transmembrane domain is referred to as the N-terminal domain (NTER). The role of these residues (Ala168-Gly550) has been investigated extensively. As with the 28-158 fragment, it is possible to express this domain as a recombinant protein that is both soluble and folded [28]. This recombinant protein is able to bind agonists and antagonists in a comparable fashion to the native receptor and is a candidate for X-ray or NMR-derived structure determination.

As mentioned above, the sequence of the NTERs of class C 7TM receptors is significantly similar to that of a family of bacterial periplasmic amino acid binding proteins (LIVBP etc). The three-dimensional structures of several of these bacterial proteins have been solved by Quiocho's group in Houston [64, 65], and these structures provide a basis for modelling the three-dimensional structure of the ligand-binding, NTER of GABAB R1a (Figure 1.2). A similar exercise had already proved useful in the case of mGlu R [32].

The LIVBP (Figure 1.3) is a prolate ellipsoid with dimensions 35 Å X 40 Å X 70 Å, composed from distinct globular domains (lobes I and II). The domains are linked by three short stretches of residues which are widely spaced within the primary structure such that each domain contains contributions from both the N- and C-terminal halves of the sequence. The two lobes are similar in structure and composed of a seven stranded β-sheet flanked by α-helices. The linking sequences form the bottom of a cleft between the lobes which has a depth of about 15 Å and a base of 14 Å X 16 Å. The leucine binding site is in a crevice to the side of the cleft, within the lobe that contains the N-terminal residue (lobe I). L-leucine is held in place by hydrogen-bonding of its -NH3+ and -COO- group (with the α -CO of Ala~100/-OH of Thr~102, and the -OH of Ser~79, respectively), while its side chain lies in a depression lined with the non-polar residues (Phe~276, Tyr~18 and Leu~77).

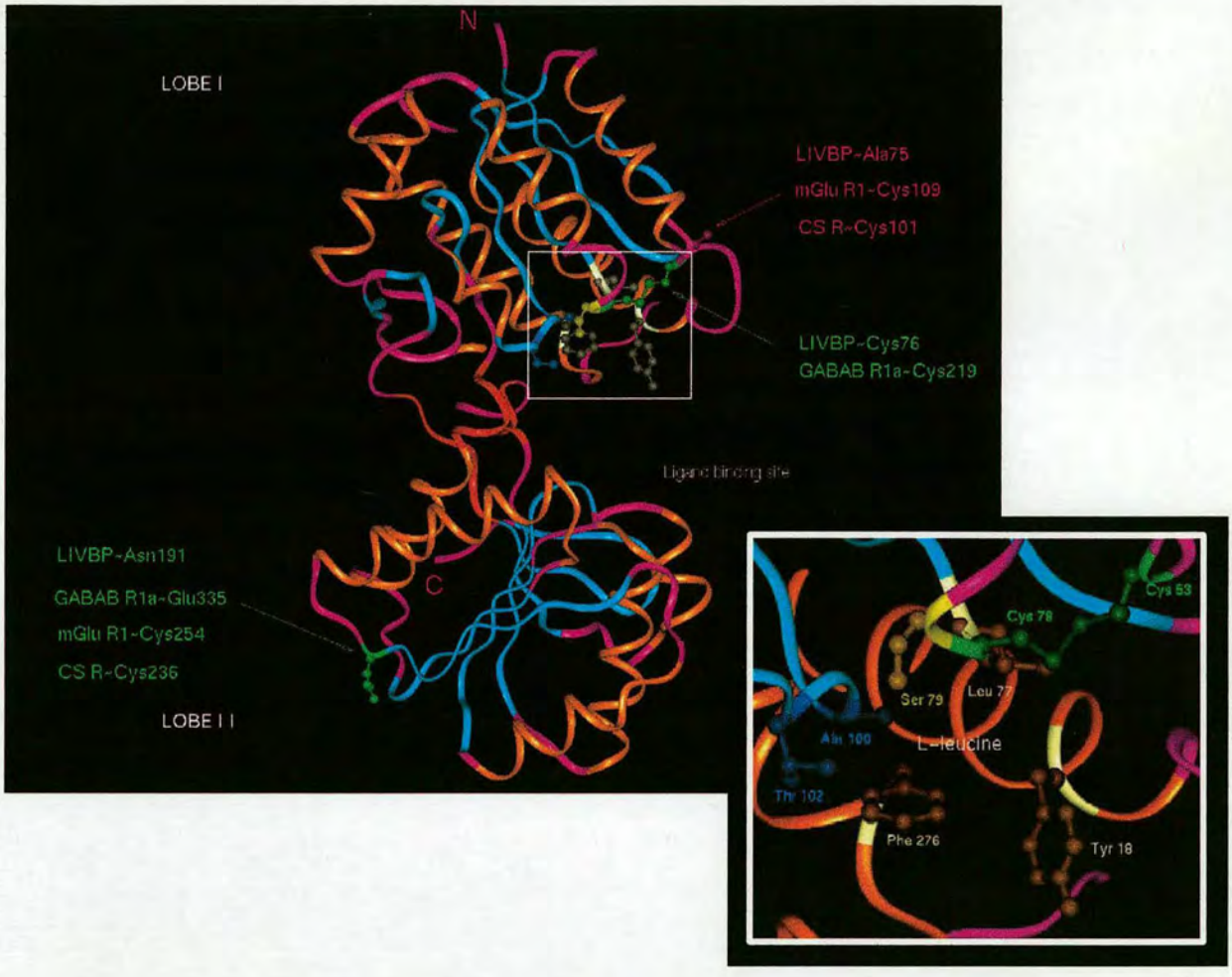



**Figure 1.3: Crystal structure of the Leucine/Isoleucine/Valine-binding protein from *E. coli* [64]**

Colored, are the secondary structure predictions for the $GABA_B$ R1 N-terminal domain [33] (Figure 1.2), that is sufficient to specify agonist and antagonist binding [28].

A plausible model of the NTER of $GABA_B$ R1a/b was built using the programme MODELER [33]. The model resulted in several interesting observations. There are five cysteines within the NTER of $GABA_B$ R1b (compared to 17 in the NTER of mGlu R), and in the model these are seen to pair up such that Cys~219 and Cys~245 could form a disulphide bond, as could Cys~375 and Cys~409; the remaining Cys, 187, is buried in lobe 1 and not in the proximity of the other Cys residues. The model also supplied predictions of the residues of $GABA_B$ R1a/b likely to be involved in binding of GABA, assuming that these are equivalent to the residues of LIVBP involved in binding leucine. For example, Ser~246 and Ser~269 of GABA B R1a/b are equivalent to Ser~79 and Thr~102 of LIVBP.

S246A, and another mutant, Y470A suppressed ligand binding altogether [33, 66]. Ser~246 (Ser~79 in LIVBP) is almost certainly involved in a direct interaction with GABA, and it is pertinent that the equivalent residue in mGlu R (Ser~165) is also critical for ligand-binding. The mutation S269A exerted a small selective influence on ligand binding. The equivalent residue, Thr~102 of LIVBP, is critical for H-bonding to the $\alpha$-$NH_3^+$ of leucine, a moiety which is lacking in GABA, so the lack of a major effect for this mutation may be rationalised. The equivalent residue in mGlu R (Thr~188) is functionally critical, as expected. Interestingly, the equivalent residue in $GABA_B$ R2 of $GABA_B$ R1 S269 is a proline, and therefore could explain the lack of GABA potency at this subunit.

Hence mutagenesis studies supported the reliability of the model. This encouraged speculation that amongst these proteins there might be parallels between the molecular mechanisms underlying function. In the case of the bacterial periplasmic amino acid binding proteins, crystallographic evidence has accumulated for a ligand-induced hinge-bending motion. This so-called "venus flytrap mechanism" in which binding of the ligand is followed by a conformational change that buries the ligand between the two domains, might exist also in $GABA_B$ R1 (and other class C receptors). Indeed, two mutants, S247A in lobe I and Q312A in lobe II were proposed to manifest their functional effects - a small loss of antagonist affinity accompanied by a gain in agonist potency - by influencing directly the hinge-bending motion.

Because $GABA_B$ receptors share sequence similarity with mGluRs and CaSR, two types of receptors that are sensitive to $Ca^{2+}$, a possible regulation by $Ca^{2+}$ ions was investigated [67]. A large effect of $Ca^{2+}$ was observed on GABA-binding, where the affinity is decreased by a factor >10 in the absence of $Ca^{2+}$ but not in the case of binding to baclofen and two other antagonists. The potency for $Ca^{2+}$ in regulating GABA affinity was 37 $\mu$M. Investigation of the $Ca^{2+}$ sensitivity of a series of $GABA_B$ R1

mutants identified residue S269 to be critical; the S269A mutation decreased affinity for GABA but abolished $Ca^{2+}$- sensitivity [67].

More recent mutagenesis studies from Jean-Philippe Pin's team [66], suggest that S246 and D471 located within lobe I form hydrogen bonds and a salt bridge with the carboxylic and amino groups of GABA, respectively, demonstrating the pivotal role of lobe I in agonist binding. Interestingly their data also suggest that residue Y366 in lobe II interacts with the agonists in the closed state, Y366A mutation convert agonist baclofen into an antagonist. Taken together these results strongly suggest that a closed state of the binding domain is required for receptor activation.

### 1.1.6.3 Allosteric interactions between $GABA_B$ R1 and $GABA_B$ R2 subunits

The recent crystal structure of the extracellular ligand binding domain region of mGluR1 (Figure 1.4.1) offers an insight into the activation mode of the Family 3 G protein coupled receptors [68]. Three different crystal structures were determined, two unliganded forms, called free form I and II, and a complex form with glutamate. All three structures showed disulphide-linked homodimers, where each monomer is structurally related to LIVBP, and has the two expected globular domains (lobes I and II) forming a 'clamshell'-like shape. The intermonomer disulphide bond between Cys 140 of each monomer is located within a disordered fragment and lobe I provides the dimer interfaces in all three forms. mGlu R1 shows an "open-closed" conformation (similar to LIVBP); the complex form has one monomer in the open conformation and the other in the closed conformation. Each monomer binds glutamate at similar sites in lobe I, and all ligand-binding residues correspond to the ones previously identified by mutagenesis studies. The monomers differ in the relative orientation of lobe I and lobe II. In contrast to the complex form, the free form I consist of two open conformers while the free form II maintains the same "open-closed" conformation as the complex form despite the absence of ligand. Comparison of free form I and complex form revealed a 70° rotation on the dimer interfaces of lobe I but which is uncoupled to the open-closed conformational change.

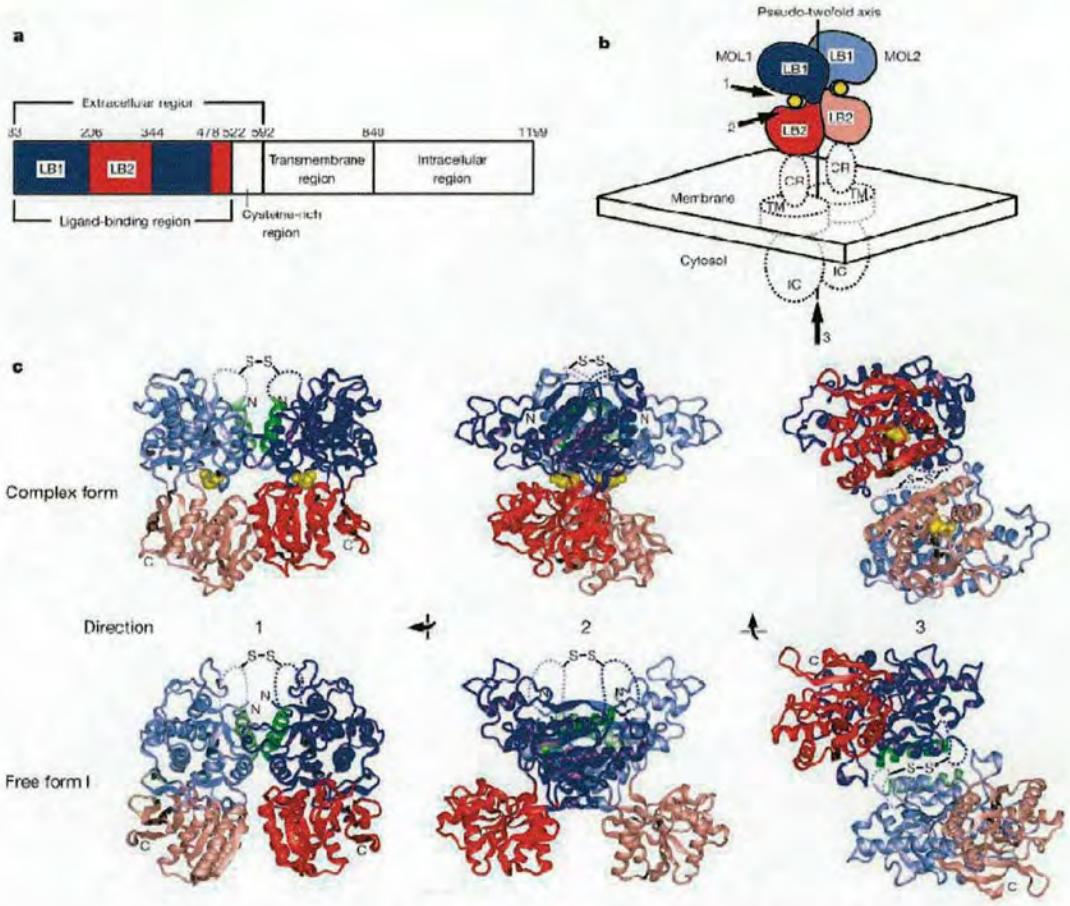



**Figure 1.4.1: Crystal structures of mGlu R1 in the liganded and unliganded states.** ***Taken from reference*** **[68].** **(a),** Diagram of full-length mGlu R1. Functional regions are boxed. The Lobe I (LB1) and Lobe II (LB2) domains, which constitute a ligand-binding region, are coloured blue and red, respectively. Numerical positions of amino-acid residues are indicated according to the primary amino-acid sequence. **(b),** Drawing of the spatial arrangements of mGluR1 domains. The molecule 1 (MOL1) and molecule 2 (MOL2) molecules of the dimer are distinguished by dark and light colouring, respectively. The ligand, glutamate, is shown as yellow spheres. CR, cysteine-rich region; IC, intracellular region; TM, transmembrane region. **(c),** Three orthogonal views of the dimer structures of the complex form and the free form I. The perspectives are indicated by arrows in b. Bound glutamate molecules are shown as yellow space-filling models. Disordered regions with a potential interprotomer disulphide bridge are indicated by dotted lines. The B helix, constituting the dimer interface between subunits, is coloured green.

The two distinct intermonomer orientations of Lobe I were designated as the 'Resting' state ('R' conformation) and the 'Active' state ('A' conformation) for the free form I and the complex form / free form II respectively (Figure 1.4.2). These findings suggest that the unliganded protein undergoes a dynamic change between an open and a closed conformation, and that the ligand stabilizes the closed conformation.

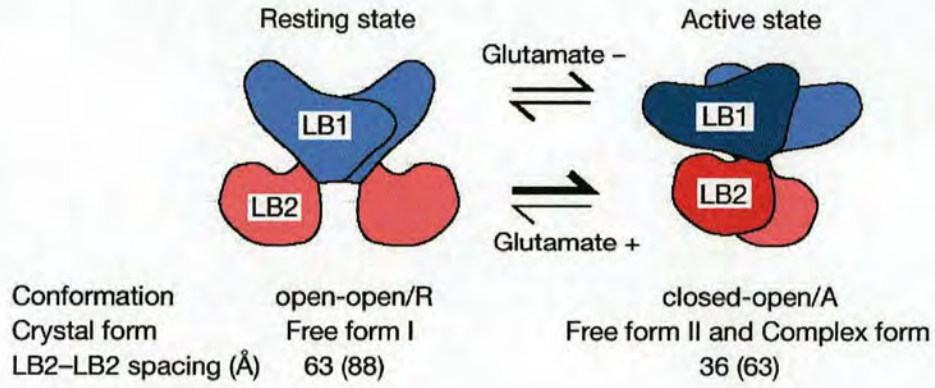



**Figure 1.4.2: Diagram of the two states of the m1-LBR in dynamic equilibrium.** ***Taken from reference*** **[68].** Open and closed conformations of mGlu R1 are distinguished by light and dark colouring,respectively. Lobe II-Lobe II (LB2±LB2) spacings are shown between the centres of masses or C termini (in parentheses) of the two LB2 domains in each conformer.

Kunishima et al [68] propose that unidentified conformations of the mGluR binding-domain in various open-closed combinations might be present in solution and that glutamate binding could therefore change the equilibrium between multiple conformers to activate the receptor. A rearrangement in the dimer will probably cause a change in the spatial arrangement of the two lobes II and as a consequence induce the approach of two transmembranes regions linked to intracellular regions thereby activating G proteins, as previously described in the case of the dimeric erythropoietin receptor [69, 70]. Interestingly, further modelling studies on diverse combinations of the homodimer (closed-open/R, closed-closed/R, open-open/A, closed-closed/A) shows two families of lobe II – lobe II spacings determined exclusively by the intermonomer R-A conformations. This supports the idea that the two arrangements of the lobes II could indeed represent the inactive and active form of the receptor.

In a more recent study, Galvez et al [71] shows that the $GABA_B$ receptor requires the presence of both subunits R1 and R2 for efficient coupling, the only known example of such phenomenon among GPCRs. Using chimeric receptors, it was shown that the helical domain (the seven α-helices, loops and the C-terminus) of $GABA_B$ R2 contains the molecular determinants required for G-protein signalling. Reciprocally, the $GABA_B$ R1 heptahelical domain improves coupling efficacy. Taken together these data strongly suggest that multiple allosteric interactions between the two subunits are required for proper functioning of the receptor. This is in agreement with Kunishima's hypothesis [68] where the closure of a single extracellular domain is sufficient to bring about the proposed active state of the dimer [71].

### 1.1.7 Developmental regulation

The distinct sizes of the $GABA_B$ R1a and R1b splice variants, together with specific antibodies capable of distinguishing R1a, R1b and R2 on Western blots, allows investigation of the developmental regulation of these $GABA_B$ receptor subunits in the rat central nervous system. Using an antibody (Ab174.1) raised against a C-terminal stretch of 59 amino acids common to both R1a and R1b, Malitschek et al. [72], examined $GABA_B$ R1 expression in rats from postnatal days two to 28. The summed levels of R1a and R1b decrease in the cortex, the cerebellum and most dramatically in the spinal cord between postnatal day seven and adulthood. In these regions, $GABA_B$ R1a levels appear significantly higher than R1b between days two and 14 although no quantitation of the Western blot immunoreactivity was included in this report. As $GABA_B$ R1a levels decrease over this time period there appears to be a slight increase in R1b levels. After day 28, some reduction in expression of both R1a and R1b is observed in the midbrain but not in the striatum, brainstem or hippocampus.

Using the photo-affinity label, [$^{125}$I]CGP64213, the developmental changes in the affinity of rat cortical membrane $GABA_B$ receptors for L-baclofen were examined in competition-binding assays. The potency for L-baclofen increases about ten-fold between postnatal days one and 60. Furthermore, SDS gel analysis of the cross-linked membranes demonstrate that the R1a and R1b isoforms do not differ in their affinity for L-baclofen between day four and adult.

The issue of the developmental regulation of $GABA_B$ R1a and R1b was further examined with another antibody (gb1a,b922-944) prepared against a synthetic peptide corresponding to a

sequence (922-944) common to both R1a and R1b [69]. It is expected that this antibody (like Ab 174.1) would have identical affinity for both R1a and R1b on Western blots and thus, following separation of the two isoforms on SDS gels, it could be used to quantitate expression levels. On the basis of the immunoreactivity observed in Western blots, it was shown that R1a levels are highest in rat brain at birth through postnatal day five during which time, R1b levels are about five-fold lower. R1b levels then increase about three-fold from day five to day ten but decrease again between day ten and adulthood to a level of about twice that seen at birth. R1a levels decline steadily after day five and in the adult rat brain there appears to be about twice as much R1b immunoreactivity as R1a.

The antibody (gb1a,b922-944) was used to examine the regional distribution of $GABA_B$ R1a and R1b in membranes prepared from various regions of the adult rat brain. R1b levels were highest in the cortex, thalamus and cerebellum where R1a levels ranged from 50-60% of R1b levels. In the striatum and olfactory bulb regions, R1a levels exceeded R1b levels but the immunoreactivity of R1a in these regions was only about 20% of that seen for R1b in the adult cerebellum.

A similar Western blot analysis with an antibody (AbC22) specific for R2 revealed that $GABA_B$ R2 is expressed at high levels in the cortex and cerebellum throughout postnatal development. In the cortex, $GABA_B$ R2 levels increase between day two and 28 with a slight apparent decrease in the adult (60 days) [15]. $GABA_B$ R2 levels in the cerebellum appear highest at four days, the earliest time point examined, and decline slightly after that. In contrast, R2 levels are detectable in spinal cord from postnatal days seven (the earliest time examined) to 14 but decline sharply thereafter. It is of interest from the perspective of potential heterodimer configurations, that while R2 levels are still increasing in the cortex between day two and 28 [15], the combined levels of R1a and R1b in the cortex decrease somewhat over this same period [72].

Recently, a detailed study of the developmental regulation of other $GABA_B$ subtypes showed that the expression of the $GABA_B$ R1c subunit was up-regulated in foetal brain, suggesting a role for this variant in neural development [49].

### 1.1.8 Receptor heterogeneity

The cloned $GABA_B$ receptors do not reproduce the pharmacological diversity of native receptors [73]. Such diversity could arise from diverse sources: differential distribution, isoforms, posttranslational modification of the $GABA_B$ subunits, and interaction of the subunits with other as yet unidentified subunits or chaperone molecules. Alternatively the differences in the potency of a particular drug in functional studies could relate to differences in the neuronal effector systems and not necessarily to the existence of numerous subtypes. Also the location of $GABA_B$ receptors at particular subcellular sites might ensure their interactions with selective effector systems and thereby display particular pharmacological actions. In this respect, sucrose gradient centrifugation analysis of deoxycholate extracts of rat brain membranes identified a discrete population of heterodimer, which has an apparent molecular weight in excess of 250 kDa [43]. This larger material might consist of complexes between the heterodimer and an as yet unidentified associated synaptic protein(s).

The lack of expression of $GABA_B$ R2 mRNA in peripheral tissues [49], also suggests that many functional $GABA_B$ receptors (as shown by previous studies [4, 21, 73]) in these tissues result from the association of $GABA_B$ R1 with yet unidentified $GABA_B$ subunits or associated chaperones proteins. These findings were confirmed at the protein level in uterus and spleen by western blotting, two types of tissues were $GABA_B$ R1 protein is present at a high level [49].

Recently, three $GABA_B$ subunits were identified from *Drosophila melanogaster* [74], while the first two subunits show high sequence identity to mammalian $GABA_BR1$ and $GABA_BR2$ , the third subunit does not relate to any known mammalian counter part and therefore could be insect-specific or simply related to yet unidentified subunit(s) in mammals.

## 1.2 The extracellular matrix protein family-fibulins.

Our collaborator, Dr. Julia White ( GlaxoSmithKline – Stevenage – UK ) has used a YTH screen against the N-terminal CP-modules of $GABA_B$ R1a to look for potential interacting protein partners. All five known variants of fibulin were recovered. The fibulins are an emerging family of secreted glycoproteins, designated fibulin-1-5. Fibulins are part of extracellular matrix structures such as connective tissue fibers, basement membranes, and blood clots and are present in the brain. Further screenings have identified the C-terminal domain that is common to all five variants to interact with the first CP-module of $GABA_B$ R1a. The YTH system has been extensively used to screen for intracellular interactions, but it is only recently that such system was applied to extracellular interactions [75]. A growing number of extracellular interactions have been detected using YTH [76-79], despite concern about the lack of disulfide bonds formation and the absence of post-translational modifications (nuclear environment). Among these extracellular interactions, several involved extracellular matrix proteins, most which contain disulfide bonds. In particular, examples of interactions involving EGF containing ECM proteins were first identified using YTH [80, 81], a protein motif that contain three disulfide bonds. Although uncertain, the YTH interaction between the first CP-module of $GABA_B$ R1a, and the C-terminal domain of the fibulin family represents the only clue that could lead to an understanding of the role of the CP-modules in $GABA_B$ R1 variants. To date, no report has been made on the function of the CP-modules in $GABA_B$ R1 variants.

Fibulins could influence $GABA_B$ receptor biology in many ways. $GABA_B$ R1a and $GABA_B$ R1b show profound differences in neuronal localisation as well as differences in expression throughout development. Moreover, not all the reported pharmacology attributed to the $GABA_B$ receptors can be explained simply by heterodimerization. A further possibility is that the fibulins could mediate protein contacts between $GABA_B$ receptors and other extracellular proteins, either within the same cell to achieve a clustering effect that could also involve an effector (ion channel for example), or with a second cell, to provide cell-cell contacts. Such clustering roles are well described in the case of intracellular linking proteins such as the PDZ-containing proteins [37, 82], and extracellular matrix proteins could be likewise involved. Occurrence of receptor clustering by ECM protein in CNS synapses has been reported; NARP (neuronal activity-regulated pentaxin) [83] is an ECM protein that can induce clustering of AMPA-type glutamate receptors [84, 85].

### 1.2.1 The extracellular matrix

The extracellular matrix (ECM) can be viewed simplistically as a three dimensional network of interacting molecules that surrounds cells. The ECM contains both collagens and noncollagenous glycoproteins, such as proteoglycans, a group of glycoproteins that have covalently bound sulfated glycosaminoglycan chains (aggrecan, versican, perlecan, brevican, neurocan, etc) and multiadhesive matrix proteins, which bind cell-surface adhesion receptors. Multiadhesive matrix proteins are long and flexible molecules that contain different binding regions for various collagen types, other matrix proteins, polysaccharides, cell surface proteins, growth factors, and hormones. The laminins are a family of cross-shaped multiadhesive proteins and an essential component of the basal lamina, a thin sheet-like network of ECM proteins where most epithelial cells rest. Laminins have a large heterotrimeric structure and are able to bind type IV collagen, sulfated lipids, integrins and heparan sulfate. Fibronectin is another major class of multiadhesive matrix proteins. Fibronectin is a large multidomain glycoprotein that attaches cells to collagen and thereby regulates the shape of the cells and the organisation of the cytoskeleton. Fibronectins are disulphide-linked dimers, each chain having similar sequences and composed of three types of repeating motifs. Fibronectin also binds a wide range of ECM components such as fibulins, integrins, heparan sulfate, and fibrin. One further important component of the ECM, is hyaluronan, a large polysaccharide that forms a hydrated gel and gives the ECM its resistance to compression. The composition and organisation of ECM in tissues is variable and its interaction with cells shape differential patterns of gene expression and cellular behaviours including guidance of migrating cells during development. Fibulins are rod-shaped proteins that interact with many other ECM proteins such as fibronectin, laminins, nidogen, perlecan, fibrillin and fibrinogen. Fibulins contain three domains (Figure 1.5); one N-terminal domain (domain I), a central array of EGF/EGF-like calcium binding modules that differs in length depending on subtypes (domain II), and one C-terminal domain (domain III) that represents a novel sequence motif of approximately 120 residues, and is known as the fibulin type module. Subtypes differ in their domain I composition, and the number of EGF/EGF like modules in domain II. Fibulin-1 and fibulin-2 are best characterised. They have been identified as basement membrane and microfibrillar proteins [86] with a broad range of binding range specificities to other extracellular ligands. Several of these interactions involve domain II and are therefore calcium-dependent.

### 1.2.2 Fibulin-1

Fibulin-1 was the first member of the fibulin family to be isolated [87]. Alternative splicing of fibulin-1 results in four different transcripts, fibulin–1 A-D (62-77 kDa), which differ in C-terminus length. Only the C-termini of fibulins 1C and 1D have the fibulin type module, although the fibulin 1D fibulin type module is 20 residues longer than the fibulin type module of fibulin 1C [88]. Domain I of fibulin-1 contain three anaphylatoxin-like modules sometimes known confusingly as complement repeats and domain II comprise an array of nine EGF-like repeats. Fibulin-1 has been shown to bind the extracellular matrix proteins fibronectin, nidogen [89], laminin and the coagulation protein fibrinogen to promote platelet adhesion (thrombosis). In addition, fibulin-1 is capable of self-association [90]. Interestingly, both self-association and fibronectin-binding sites are identical and map to EGF modules five and six in domain II [91]. Fibulin –1 is also a ligand for the C-type lectin domains of aggrecan and versican [92], two proteogylcans that can be found in the brain, but it does not bind brain-specific brevican and neurocan. Recently, binding of fibulin-1 to the amino-terminal head of β-amyloid precursor protein (APP) was demonstrated using YTH and confirmed by GST-pull down assays. The APP-binding site maps to EGF modules 3-7 in domain II. Furthermore, fibulin-1 was found to be produced principally by neurons and to inhibit secreted APP-mediated neural stem cell proliferation [81].

### 1.2.3 Fibulin-2

Fibulin-2 was first identified by cDNA cloning and sequencing [93]. Fibulin-2 is a 175-kDa protein that forms a disulphide-linked homodimer in which the subunits are arranged in a head-to-tail manner with their disulphide bond connection through domain I and located in the center of the dimer [86]. Fibulin-2 has a larger domain I when compared to fibulin-1, which contains a cyteine-rich region followed by a cysteine-free region. Fibulin-2 is found in endocardial tissues, neural and bone forming tissues. *In vitro* binding studies have demonstrated affinities for basement membrane components nidogen [94], perlecan, collagen IV, fibronectin, and fibrillin that in most cases were dependent on calcium [95]. The proteoglycans aggrecan, versican and brevican form networks with fibulin-2 through their lectin domain and their interaction sites were mapped to EGF modules 3 and 4 in domain II [96]. Fibulin-2 also interacts with laminins, and fibulin-2 functions in assembling the

laminin network in the basal lamina at the dermal-epidermal junction by bridging laminin-1 and laminin-5 with other matrix proteins [97].

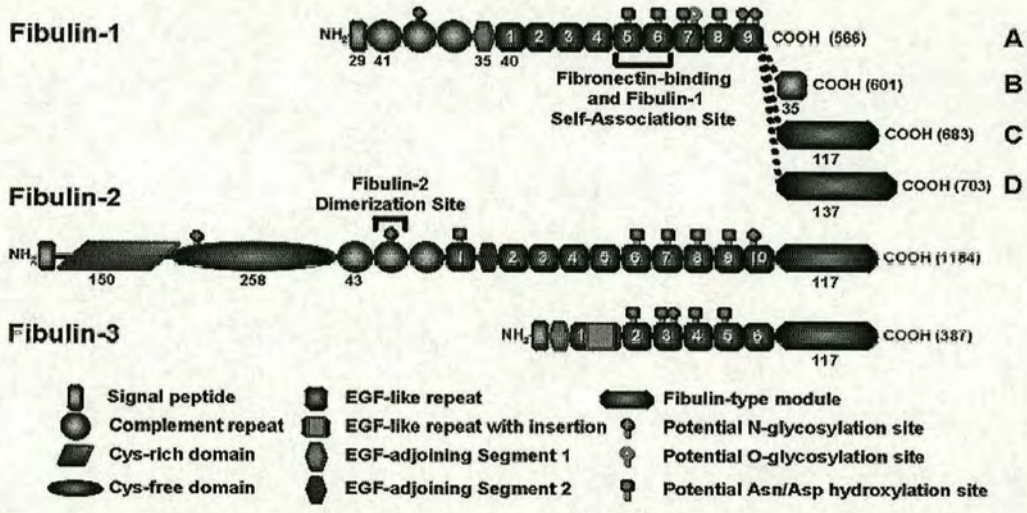



**Figure 1.5: Schematic diagram depicting the domain organization of fibulin-1 (forms A-D), -2 and –3 - *Taken from Argraves WS - http://ef.wustl.edu/cartoons/fibulins.gif***

### 1.2.4 Fibulin-3 and -4

Fibulin-3 (57 kDa) and fibulin-4 (54 kDa) are of similar composition: a short domain I that consist of a divergent EGF-like module, an array of five EGF-like modules and the fibulin-type module [98]. Both fibulin-3 and fibulin-4 are monomeric and are expressed in several human tissues to a variable extent. Nothern blot ananlysis showed strong expression of fibulin-3 and fibulin-4 in heart and weak expression in brain [99]. Electron microscopy indicated a short rod-like structure for fibulin-4.

### 1.2.5 Fibulin-5

Fibulin-5 is a 55 kDa protein similar in structure to fibulin-3 and -4, and was first identified as EVEC or DANCE proteins [100, 101]. Fibulin-5 has also a short domain I made of one modified EGF modules that contains a RGD sequence. This motif is recognised by many integrins, a large family of

cell surface receptors that mediate cell-cell and cell-matrix adhesion. Fibulin-5 domains II and III are homologous to those of fibulin-3 and -4. Northern blot analysis also shows strong expression in heart and weak expression in brain.

Recently, studies on fibulin-5 null mice showed that fibulin-5 is an essential element for elastic fibre development, which acts as a scaffold protein that organizes and links elastic fibres to cells [102, 103].

### 1.2.6 The fibulin type module

Fibulins are distantly related to fibrillins with which they share the C-terminal fibulin type module (Figure 1.6). A sequence comparison of their C-terminal domains revealed 23% identity between fibulin-3/4 and Fibulin-1C/2, and 11-16% identity with the corresponding region of fibrillin-1/2 [98]. The alignment shows two invariant cysteine residues and one proline at the N-terminus, and the consensus sequence IXXGNXXXXF in the central portion. Seven proteins share this recently described sequence motif. Fibulin-1 has different subtypes (A-D) that differ in their domain III. The C-terminus of fibulin-1B C-terminus has only 35 extra residues when compared to fibulin-1A, which does not contain the fibulin-type module. Fibulin-1C/D has the complete sequence motif. Several binding sites for fibulin ligands involve the fibulin-type module. For example, nidogen binds fibulin1C/D at domain II/III, and deletion of the fibulin-type module induced a 30-fold reduction in binding [89]. Interestingly nidogen-binding to variant C is 100-fold stronger than for variant D [104]. The NOVH protein also binds fibulin-1C at domain II/III; NOVH belongs to the CCN family of growth factors [105]. Although the biological function of fibulin-type module remains to be established, its presence in fibrillins raises interesting comparisons. Furthermore it has been shown that fibrillin-1 and fibulin-2 are co-localised in some tissues (extra-CNS) and binding of fibulin-2 to the N-terminus of fibrillin-1 has been demonstrated [106]. Fibrillins are the major constituents of extracellular microfibrils. The two fibrillin isoforms are large glycoproteins (~350 kDa) with a complex multidomain structure, including 47 EGF-like modules [107].

How fibrillin molecules assemble into microfibrils is not known. Initial investigations suggest that microfibrils are stabilised by disulphide bonds, and the use of reducing agents abolishes their ultrastructure [108]. The N- and C-terminal domains of fibrillins display respectively four and two

cysteine residues. All four cysteine residues within the N-terminal domain are involved in intrachain disulphide bonds.

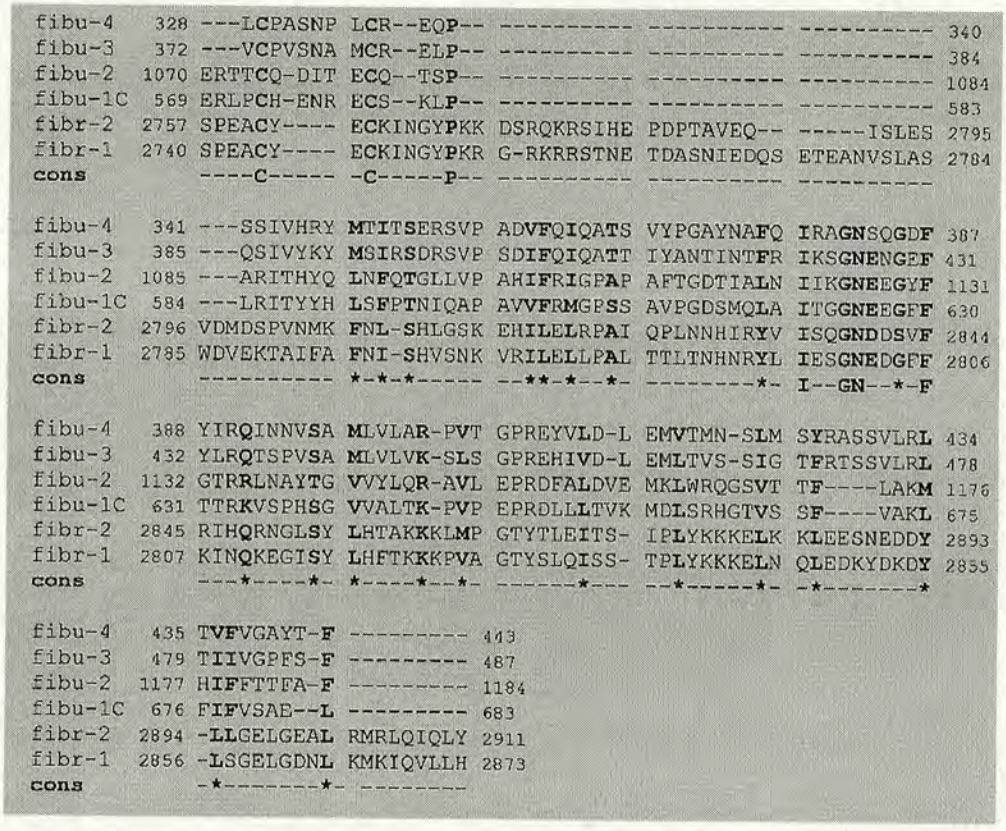



**Figure 1.6: Sequence alignments of the C-terminal domains of fibulin and fibrillin isoforms**

***Taken from reference*** **[98].** Identical or similar amino acids in at least five sequences are shown in a in bold. The consensus sequence (cons) indicates similar (asterisk) or identical amino acids in all sequences. Hyphens in the sequences indicate gaps introduced for optimal alignment. All sequences compared were of human origin.

A single non-conserved cysteine residue was found in fibrillin-1 and could therefore be available for intermolecular disulphide bond formation [108]. No information is yet available about the form of the two cysteine residues in the fibulin-type module. Fibrillins undergo proteolytic cleavage of both the leader sequence and their fibulin-type module. Cleavage of the fibulin-type module has been proposed to occur after arginine residue 2731 following the multi-basic sequence KRGRKR, and

results in the loss of the sequence that follows the conserved two cysteine residues and the proline residue (Figure 1.6) [109]. A recent study confirmed that the region is processed intracellularly, before fibrillin-1 secretion, by an enzyme that resides in the early secretory pathway. Proteases of the Furin family have affinity for such tetrabasic sequences, and it has been suggested that it is these enzymes that are responsible for this early processing. Interestingly, intermolecular disulphide bond formation occurs within the first hours after synthesis and secretion of monomeric fibrillin-1, and could indicate that C-terminal processing is involved in assembly of fibrillins. It is tempting to speculate about fibulin-type module function. Ritty et al have proposed that the fibulin-type module helps the association of monomers into multimers, and once that function has been accomplished, the domain is processed. Assembly into multimers before secretion has already been described for other matrix proteins such as laminin or tenascin [110]. Disulphide bond exchange has also been proposed to explain this multimerisation process [107], and it could be that cleavage of fibulin-type module is required for such a mechanism.

# CHAPTER TWO: NUCLEAR MAGNETIC RESONANCE THEORY

## 2.1 Basic principles of NMR

### 2.1.1 Nuclei in magnetic field

Atomic nuclei are composed of protons and neutrons. Each of these particles has a spin angular momentum ($\vec{I}$). The magnitude of this vector quantity is expressed in units of spin quantum number (I). Protons and neutrons have I value of ½. Nuclei are also characterised by a spin quantum number and nuclei I value may have 2I+1 discrete values in integral steps from –I to +I (I = -I, -I+1,….., +I), where I = n/2, $\|\vec{I}\| = [I(I+1)]^{1/2}\hbar$ , n is an integer, $\hbar (= h/2\pi)$, and h is Planck's constant [111-113]. Only the ground states of nuclear spins are populated significantly. In general, there are no simple rules for which of the many possible states is the ground state, and such a state is best regarded as an empirical property of each isotope. If I>0, the nucleus has a magnetic dipole moment, $\vec{\mu}$, and can be seen as a charged, spinning particle producing a small circular current. $\vec{\mu}$ is directly proportional to $\vec{I}$: $\vec{\mu} = \gamma\vec{I}$. The proportionality constant $\gamma$ (rad$^{-1}$ T$^{-1}$ s$^{-1}$or rad$^{-1}$ Gauss$^{-1}$) is known as the gyromagnetic ratio of the nucleus. $\gamma$ has either positive or negative value and is isotope specific.

Iz, the z component of $\vec{I}$ (here the z axis is the applied static magnetic field) is quantified: Iz = $m_I \hbar$, as is the z component of $\vec{\mu}$, $\mu_z = \gamma\ m_I \hbar$, where $m_I$ is the magnetic quantum number and may have 2I+1 values in integer steps. In the absence of an applied magnetic field, all nuclear spins are disordered, and there is no energy difference between them: they are degenerate. Since they have a magnetic moment, when we apply a strong external magnetic field ($\vec{B}_0$), they take their allowed orientations. The magnetic moment of a spin ½ nucleus e.g. $^{1}H$, $^{15}N$, $^{13}C$, $^{31}P$, $^{19}F$, has only two permitted states in the static magnetic field, the $\alpha$ state which corresponds to the magnetic moment aligned parallel with the field ($m_I$=+1/2), and the $\beta$ state which corresponds to the magnetic moment

being anti-parallel to $\vec{B}_0$ ($m_I$=-1/2) [111-113]. The interaction energy of a spin state in the static magnetic field $B_0$ can be described as: $\vec{E} = -\vec{B} \bullet \vec{\mu}$. In z the direction, $E = -B_0.\mu_z = -B_0.\gamma\, m_I\, \hbar$.

For the two possible spin states of a spin ½ nucleus ($m_I = \pm$ ½), the energies are:

$$\vec{E}_\alpha = -\frac{1}{2}\gamma\hbar\vec{B}_0$$

$$\vec{E}_\beta = +\frac{1}{2}\gamma\hbar\vec{B}_0$$

The energy difference $\Delta E = E_\beta - E_\alpha = h\upsilon = \gamma\hbar B_0$ corresponds to the energy that can be absorbed or emitted by the system (Figure 2.1). Only transitions between adjacent energy level are allowed ($\Delta m \pm 1$) and the resonance condition is $\Delta E = h\upsilon = \hbar\omega_0$, where $\omega_0$ is the Larmor frequency. The $^1$H nucleus has a resonance frequency of 600 MHz in a magnetic field of ~14.1 T. For comparison, Earth's magnetic field is 50 μT.

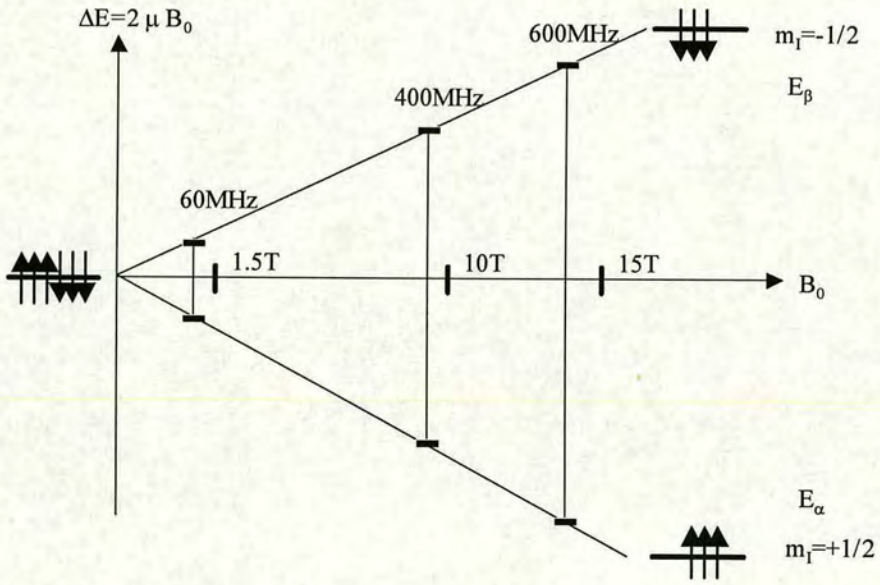



**Figure 2.1: Energy level diagram**

When placed in a static magnetic field $\vec{B}_0$, the force generated by $\vec{B}_0$ on the magnetic moment is a torque, which cause its precession about $\vec{B}_0$. The Larmor frequency can be understood as the precession frequency of the spins about the axis of the magnetic field $\vec{B}_0$, caused by the magnetic force acting up on them.

The two energy states for $^1H$ will be unequally populated, the ratio of populations being given by the Boltzmann equation, $\frac{N_\beta}{N_\alpha} = e^{-\frac{\gamma\hbar B_0}{kT}}$. The population difference is thus dependent both on the field and also on the isotope. This population difference is remarkably small ($10^{-4}$ to $10^{-5}$) and corresponds to a bulk magnetisation or macroscopic magnetic moment ($\vec{M}_0$), which is in effect, the vector sum of the magnetisation of the individual nuclear spins across the whole sample that precess around $\vec{B}_0$ (Figure 2.2).

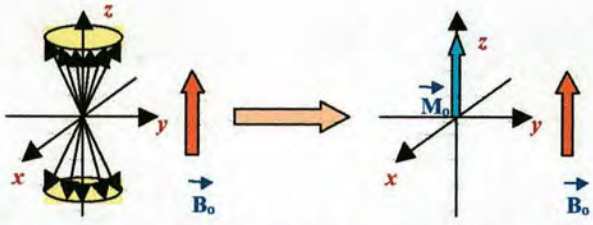



**Figure 2.2: Bulk magnetisation vector** - $\vec{M}_0$ is the vector sum of the magnetisation of the individual nuclear spins (black arrows) across the whole sample that precess around $\vec{B}_0$.

### 2.1.2 Effects of pulses

The equilibrium magnetisation is aligned along the static field $\vec{B}_0$ and represents the starting point for all NMR experiments. If the magnetisation is moved away from the equilibrium position, its xy components will induce a small oscillating current in the receiver coil placed in the xy plane. This current is amplified, digitised, and recorded as a "Free Induction decay" (FID). Fourier transformation of the FID yields an NMR spectrum.

The magnetisation ($\vec{M}$) can be tipped away from the z axis by introducing a second field $\vec{B}_1$ perpendicular to $\vec{B}_0$, $\vec{M}$ will precess around the resultant field ($\vec{B}_0 + \vec{B}_1$) and when $\vec{B}_1$ is switch off, $\vec{M}$ will no longer be aligned with $\vec{B}_0$. The $\vec{B}_1$ field is induced from a current in a coil perpendicular to $\vec{B}_0$, effectively the transmitter, which generates a linear oscillating field that is equivalent to the sum of two counter-rotating components, one being exactly in resonance with the precessing spins, i.e. its energy matches the energy difference between the $\alpha$ and $\beta$ levels. The effects of pulses on $\vec{M}$ are best described by switching from the laboratory coordinate system to one rotating "on resonance" with the spins (and $B_1$) about the z axis. The magnetisation is static in this "rotating frame" (Figure 2.3).

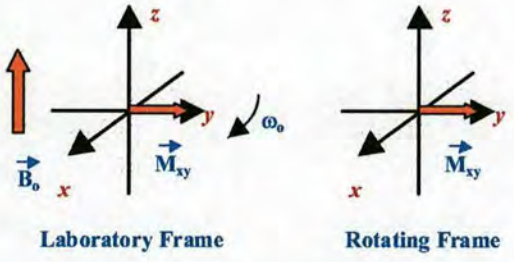



**Figure 2.3: Magnetisation vector (x, y components) in the laboratory and rotating frames**

For resonance to occur, the excitation frequency ($\nu_1$) needs to equal the nuclear precession frequency ($\nu_0$): $\nu_1 = \nu_0 = \gamma B_0 / 2\pi$. The final position of $\vec{M}$ will depend upon the length of time for which the $\vec{B}_1$ field is applied. This is commonly referred to as a radio frequency "pulse" of duration $t_p$. The flip angle $\theta$ through which $\vec{M}$ is tipped away from $\vec{B}_0$ axis is given by: $\theta = \gamma B_1 t_p$, and can be changed by varying the duration or the intensity of the radio frequency field, $\vec{B}_1$. If we measure $\vec{M}$ on the y axis (following a pulse applied along the x axis), there will be a sinusoidal variation of the signal intensity as a function of pulse duration. The signal will be a maximum for a tip angle of $\pi/2$ (Figure 2.4), null at $\pi$ when M is placed along the –z axis, a negative maximum at $3\pi/2$, and a return to a null at $2\pi$ when back on the z axis. This test is used daily on the spectrometer to calibrate pulse lengths. Note: the pulse flip convention used here is clockwise.

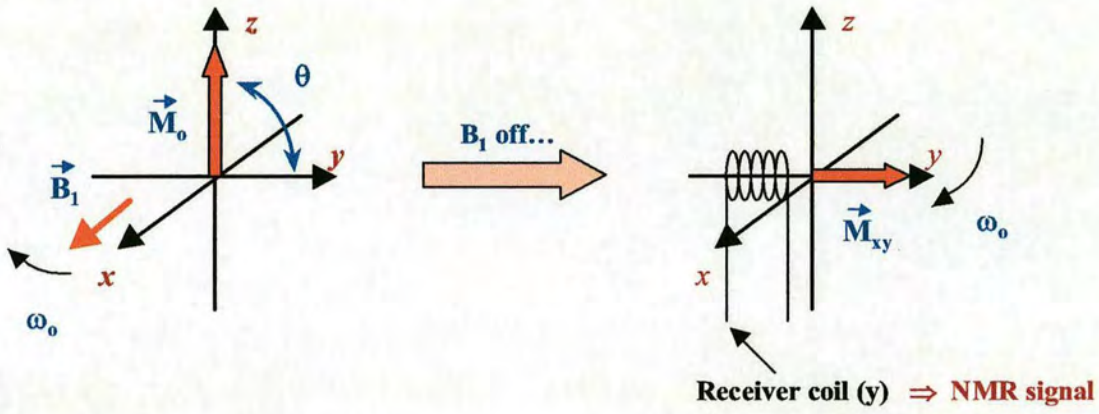



**Figure 2.4: Effect of a 90°x pulse**

When the $\vec{B}_1$ field is switched off, the transverse $\vec{M}$ (xy component or $M_{xy}$) starts precessing about the z axis with the frequency ω (in the laboratory coordinate system) under the influence of $B_0$. Thus, after a $\pi/2$ or 90° x pulse, we have: $M_y(t) = \{M_y \cos(\omega t) + M_x \sin(\omega t)\} \exp^{(-t/T2)}$, where $T_2$ is the so-called transverse relaxation. This induces a current in the receiver coil (sometimes called the probehead), here located on the y axis, which decays with time i.e. the FID (Figure 2.5). The FID is therefore composed of exponentially decaying sine and cosine signals.

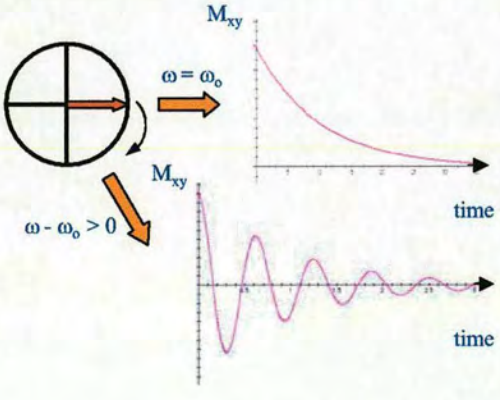



**Figure 2.5: Free induction decay**

### 2.1.3 Relaxation

The excited state of $\vec{M}$ in the xy plane is driven back to the Boltzmann equilibrium by two mechanisms:

-spin-lattice or longitudinal relaxation, $R_1$ ($=1/T_1$).

The 90° x pulse equalises the populations of the energy levels whereas a 180° x pulse inverts the population ratio, but both cause a population change, effectively a change in the energy of the spin system. In order for $\vec{M}_z$ to return to its equilibrium position, the energy absorbed from the pulse needs to be transferred to the surroundings or lattice, which in reality are the neighbouring molecules in solution. The return to $\vec{M}_z$ equilibrium position occurs with a rate constant $R_1=1/T_1$ where $T_1$ is the longitudinal relaxation time constant.

-spin-spin or transverse relaxation $R_2$ ($=1/T_2$).

Immediately after a 90° x pulse, the nuclear spins are precessing with identical phase, this condition is called phase coherence. The $T_2$ relaxation corresponds to a loss of phase coherence; this means that the precessing spins, which were bunched together, are changing randomly their phases in the xy plane causing a decay of signal with a relaxation rate $R_2=1/T_2$, where $T_2$ is the transverse relaxation time constant. $\vec{M}_{xy}$ relaxes to its equilibrium value of zero. Due to $\vec{B}_0$ being not perfectly uniform across the sample volume, some fanning out of signals occurs, which contributes to increase relaxation rates and results in additional line broadening. A symbol $T_2^*$ is used to describe the effective spin-spin relaxation.

$T_2$ determines the maximum length of a series of pulses that can be used in a particular acquisition, and for how long it is necessary to acquire the FID. $T_1$, on the other hand imposes a minimum "relaxation" delay between scans.

In the Boltzmann equilibrium, all transverse magnetisation must have disappeared so $T_2 \leq T_1$.

The $T_1$ and $T_2$ can be measured using the following pulse sequences:

Measuring $T_1$:

Inversion recovery experiment: 180° x pulse-time delay (DE)-90° x pulse-acquisition. Where the magnetisation M can be expressed as a function of time delays: $M(DE) = M_0 (1-2\exp(-DE/T_1)$.

Measuring $T_2$:

Spin echo experiment is: 90° x pulse-[time delay(DE)-180° x pulse-time delay(DE)]$_n$- -acquisition. Where the magnetisation M can be expressed as a function of time delays: $M(DE) = M_0 \exp(-2nDE/T_2)$

The spin-spin relaxation is induced by fluctuating magnetic fields at the Larmor frequency. There are two major mechanisms by which these fluctuating fields are generated: dipole-dipole interactions between different spins, and the chemical shift anisotropy. Dipole-dipole induced relaxation is the main relaxation mechanism for spins close to one another in space. The magnitude of the induced field depends on the distance between the two spins, and on the angle formed between the inter-nuclear vector and the static field. The molecule tumbles and this angle changes permanently, causes the fluctuation of the local magnetic field. The chemical shift anisotropy arises from the non-spherical distribution of electron density that will cause a fluctuating field upon rotation of the nucleus.

Both processes are connected to the movement of spins in space due to overall tumbling of the molecules or due to internal motions. Other relaxation mechanisms include relaxation arising from scalar coupling, interaction with unpaired electrons, and interaction with nuclei with a spin > 1/2 (quadrupolar relaxation).

### 2.1.4 Discrete Fourier transformation & Fast Fourier transformation

The Fourier transformation only considered continuous functions f(t). If f(t) is a **complex** time series of length **n**, then the continuous Fourier transform is replaced with a sum, where $\omega_j$ is the frequency of the $j^{th}$ point (j=0,1,2,……,n-1) :

$$F(\omega_j) = \left[ \sum_{k=0}^{n-1} f(t_k)\, e^{-i\omega_j t_k}\, dt \right] \Delta$$ where $\Delta$ is the sampling interval and $t_k = k\,\Delta$, and

k is an integer from 0 to n-1 [114].

If f(t) is periodic and we have sampled over one complete period T so f(0)=f(n), then $T = n\Delta$ and $\omega_j = j(2\pi/n\Delta)$ :

$$F(\omega_j) = \left[ \sum_{k=0}^{n-1} f(t_k)\, e^{-i2\pi jk/n} \right] \Delta$$

and the discrete inverse FT is $f(t_k) = \left[ 1/n \sum_{k=0}^{n-1} F(\omega_j)\, e^{i2\pi jk/n} \right] 1/\Delta$

If $\Delta$ is one: $F(\omega_j) = \sum_{k=0}^{n-1} f(t_k)\, e^{-i2\pi jk/n}$

and the discrete inverse FT is : $f(t_k) = 1/n \sum_{k=0}^{n-1} F(\omega_j)\, e^{i2\pi jk/n}$

When using the discrete Fourier transformation, the frequency spectrum for positive frequencies is reflected in negative frequencies, there is an "alias" or "Nyquist pair " of the transform (Figure 2.6). The reflection occurs at the mid-point of the frequency axis. Usually, the spectral region 0 to $+\nu_{max}$ and 0 to $-\nu_{max}$ are "swapped" and the frequency axis reversed to display a spectrum with $+\nu_{max}$ to $-\nu_{max}$ running from left to right. Recall that the frequency is given here by $\omega_j = j(2\pi/n\Delta)$ so the critical or Nyquist frequency $\omega_c$ is $\omega_c = \pi/\Delta$ ( effectively $\nu_{max}$ ). If the alias overlaps with the transform, it will be impossible to sort out which is the transform and which is the alias. This happens for a rectangular function that is too narrow and that will obviously become wider after Fourier transformation in the reciprocal space. But if we sample such that the critical frequency is equal to the upper frequency limit we want to observed, we will have completely determined f(t). If a signal

has a higher frequency than the critical frequency, then it will be "aliased " into the region from $\{-\omega_c,+\omega_c\}$.

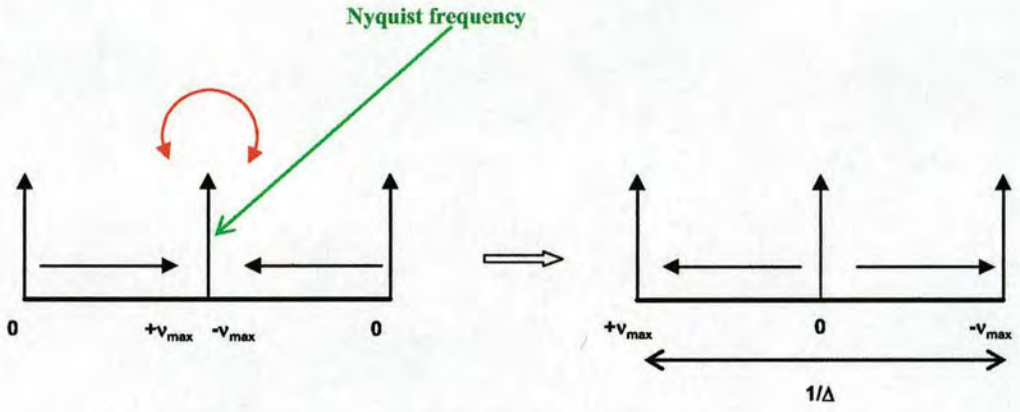



**Figure 2.6: The Nyquist pair of the Discrete FT**

We use this type of FT because we don't know f(t) from $-\infty$ to $+\infty$, but only from 0 to AQ (=length of acquisition time) and we don't acquire a continuous, but a digitised FID signal, with values known only for t = n DW, where DW is the dwell time or $\Delta$ in the equations above. The spectral window or SW in NMR corresponds to two-times the $\nu_{max}$ we want to observed i.e. from $-\nu_{max}$ to $+\nu_{max}$ so that :

$$DW=1/2\ \nu_{max}=1/sw$$

with Digital resolution = 1/n DW =1/AQ

Usually, the "carrier" sits on the reflection axis in the above description at a particular frequency, in general water, or protein NH protons resonance frequency.

In practice, the discrete Fourier transform is calculated using the fast Fourier transform (FFT) algorithm, where the discrete FT is implemented in the form of a matrix multiplication. The FFT algorithm uses symmetry properties in this matrix form of the discrete FT, to reduce the amount of computer time required for an FT by a factor of n/(log2n).

### 2.1.5 Pre and post Fourier transformation manipulations

#### 2.1.5.1 Basic manipulations

Description of the "traditional" manipulations prior FT (zero filling, and appodisation) or after FT (phase correction, and baseline correction) can be found in references [115] and [116].

#### 2.1.5.2 Maximum entropy method

One limitation of NMR is that data have to be accumulated over long periods. The time to acquire a multidimensional NMR experiment is proportional to the number of data points in the indirectly detected dimensions, causing the data for these dimensions to be almost always truncated. Although data can be zero-filled and apodised in such dimensions, apodisation technique before FT is always a "trade-off" between noise suppression and resolution enhancement. In the case of three-dimensional experiments, the truncation is severe and there are advantages in improving the resolution. Maximum entropy reconstruction is a better alternative to FT for processing severely truncated indirectly detected dimensions [116-118]. Maximum entropy is done only on indirectly detected dimensions, after FT of the directly detected dimension.

Maximum entropy reconstructs the frequency domain spectrum directly by determining the spectrum that maximises the entropy function. This function allows choosing one spectrum from the possibly infinite set of spectra which FID reproduces the experimental FID [116]. The entropy function that describes one spectrum can be viewed as inverse quantity of data it contains; therefore the chosen spectrum will be the one where the smallest amount of data was needed to reconstruct the original data set, the one that maximises the entropy function [116].

### 2.1.6 Chemical shift

The resonance frequency of a nucleus is defined as $\nu=\gamma B_0/2\pi$. However not all nuclei have the same frequency; even nuclei of the same nature do not have necessarily the same resonance frequency. In fact, the resonance frequency of a nucleus depends on the chemical "environment " that it experiences. This phenomenon is called chemical shift. It arises because the effective magnetic field B experienced by a nucleus differs slightly from the applied magnetic field $B_0$. This is due to the circulation of electrons within their atomic orbital, which creates a small local magnetic field $B_{local}$ affecting the overall field experienced by a nucleus: the nucleus is shielded from $B_0$ by its surrounding electrons. The effective field for a nucleus is $B=B_0-B_{local}=B_0(1-\sigma)$ where $\sigma$ is the shielding constant. As a result, the resonance condition becomes: $\nu=\gamma B_0(1-\sigma)/2\pi$.

If we consider the magnetisation in the xy plane, after a 90°x pulse for instance, $\vec{M}xy$ will be static along the y axis in the rotating frame, but because of the chemical shift, the frequency of excitation is slightly off-resonance and $\vec{M}xy$ precesses now at $\omega=\omega-\omega_0$ (Figure 2.7).

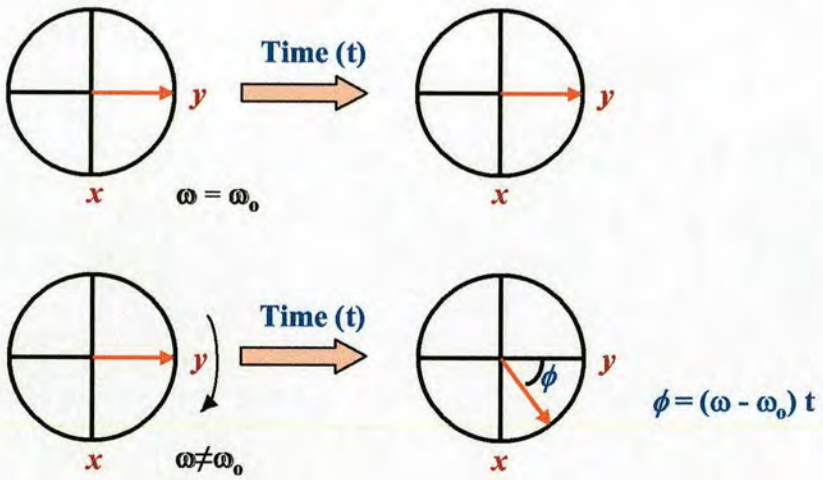



**Figure 2.7: Evolution under chemical shift**

### 2.1.7 Spin-spin coupling

The nuclear spins interact with one another by spin-spin coupling that acts through the electrons in the chemicals bonds. There are a number of factors that determine the strength of the coupling: the hybridisation of the atoms involved, bond lengths, bond angles and torsion angles etc. If two ½ spins, I and S, are coupled, spin S is aware of the presence of neighbouring spin I since the latter generates a small magnetic field at its site and reciprocally spin I is aware of spin S. This magnetic field is either positive or negative depending on the sign of a constant, the coupling constant $J_{IS}$ and on whether I is aligned or antiparallel to $B_0$. This splits the S signal into a doublet with a splitting of $J_{IS}$ Hz, reciprocally the I signal is split into a doublet at its resonance frequency (Figure 2.8). Such splitting can be removed in NMR experiments by decoupling.

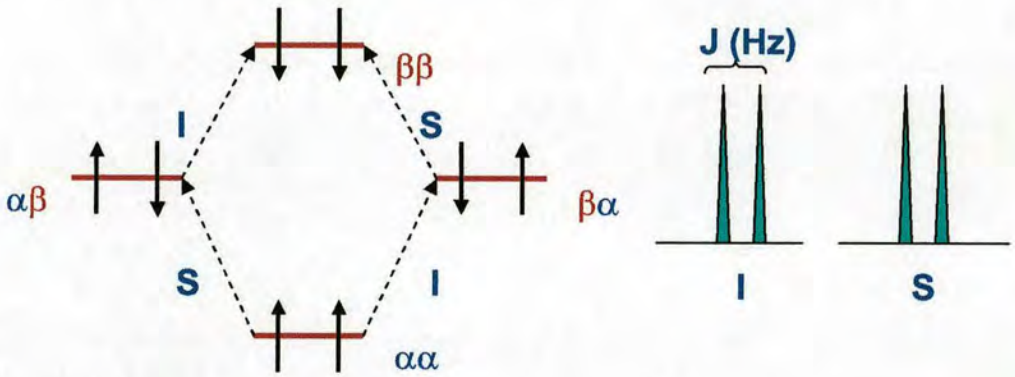



**Figure 2.8: Effects of scalar coupling on energy levels**

Spin I in the rotating frame precesses with the angular velocity $\omega_I$, and may be considered as the sum of two counter-rotating vectors (Figure 2.9). Their evolution will then depend on the magnitude of $J_{IS}$.

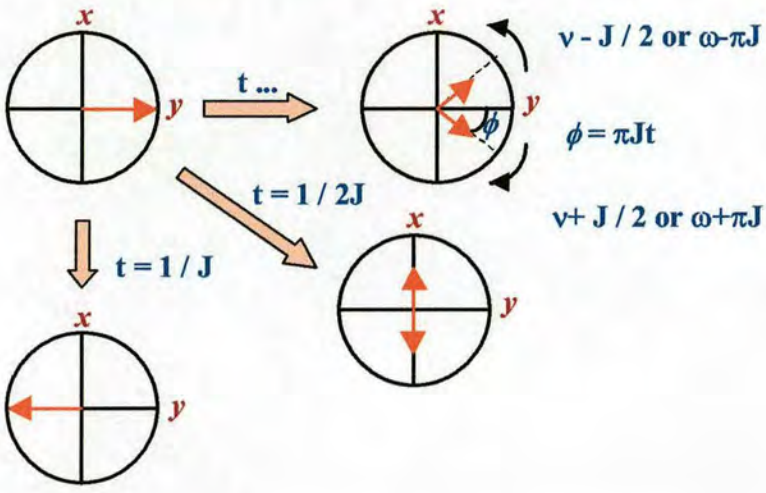



**Figure 2.9: Evolution under J coupling**

### 2.1.8 Nuclear Overhauser effect (NOE)

The Nuclear Overhauser Effect (NOE) is one of the ways in which a spin can release energy, and is therefore related to relaxation processes. The NOE is related to the exchange of energy between two spins that do not necessarily have scalar coupling but have dipolar coupling [111].

The NOE is evidenced by enhancement of certain signal in the spectrum when the populations of other spins nearby are altered. If we consider a two ½ spin I and S system energy diagram, where αα, αβ, βα, and ββ refer to states IαSβ...etc (Figure 2.10); if W represents a transition probability, at equilibrium we have $W_{1I}$ and $W_{1S}$ transitions, which represent single quantum transitions, and are related to T1 relaxation. $W_{0IS}$ and $W_{2IS}$ are double quantum transitions also known as cross-relaxation transition probabilities; these are theoretically forbidden and have lower probability, because such transitions are not directly excited by radio frequency pulses or cause directly detectable NMR signals. Nevertheless, these transitions are not forbidden in the context of relaxation. Relaxation occurs when the spins give away energy by processes that have frequencies close to $\omega=\gamma B_0$, which include movement (translation, rotation).

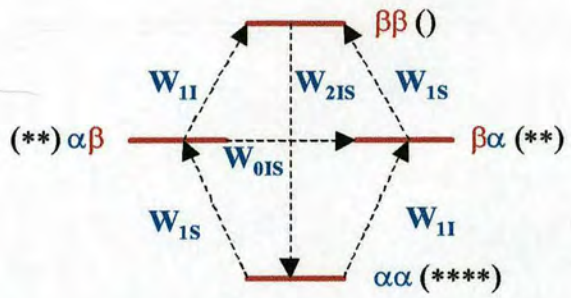



**Figure 2.10: Energy level diagram for a two spin system with $J_{IS}$=0. The probabilities W for the six allowed relaxation transitions are shown.**

If now we saturate the S transition to equalise its energy levels (Figure 2.11), $W_{1S}$ and $W_{1I}$ transitions are not observed because we have the same population in $W_{1S}$ and we did not affect spin I, but both are auto-relaxation pathways. $W_{0IS}$ and $W_{2IS}$ transitions are, in addition to the auto-relaxation, two other ways through which spin S can relax. These transitions also involve spin I, thus there will be a change in spin I populations. $W_{0IS}$ will give a negative enhancement, spin I (so far undisturbed) has deviated from its Boltzmann equilibrium towards a decrease in its α,β population difference. $W_{2IS}$ will give a positive enhancement, and spin I population difference increases.

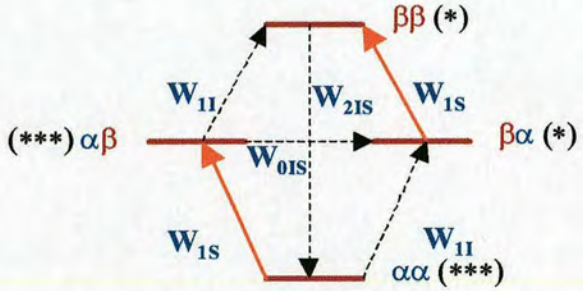



**Figure 2.11: Energy level diagram for a two spin system with $J_{IS}$=0 after saturation of spin S**

In this four level system, the different transition probabilities are:

$$W_{1I}= \frac{3}{20}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2\,\gamma_S^2\,\hbar^2\,r^{-6}\,\frac{\tau_c}{1+(\varpi_I)^2\tau_c^2}$$

$$W_{1S}= \frac{3}{20}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2\,\gamma_S^2\,\hbar^2\,r^{-6}\,\frac{\tau_c}{1+(\varpi_S)^2\tau_c^2}$$

$$W_{0IS}= \frac{1}{10}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2\,\gamma_S^2\,\hbar^2\,r^{-6}\,\frac{\tau_c}{1+(\varpi_I-\varpi_S)^2\tau_c^2}$$

$$W_{2IS}= \frac{3}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2\,\gamma_S^2\,\hbar^2\,r^{-6}\,\frac{\tau_c}{1+(\varpi_I+\varpi_S)^2\tau_c^2}$$

where $\mu_0$ is the permeability constant in a vacuum, r is the inter-nuclear distance, and $\tau_c$ is the correlation time for molecular reorientation.

The enhancement caused by the NOE can be quantified:

$$\eta = \frac{\gamma_S}{\gamma_I}\,\frac{W_{2IS} - W_{0IS}}{2W_{1I} + W_{2IS} + W_{0IS}} = \frac{\gamma_S}{\gamma_I}\,\frac{\sigma_{IS}}{\rho_{IS}}$$

where $\sigma_{IS}$ is the cross-relaxation rate between spin I and S, and it is this term that controls how quickly a NOE enhancement is transferred between the two spins; $\rho_{IS}$ is the auto-relaxation rate or dipolar longitudinal relaxation rate for relaxation of spin I.

Cross-relaxation results from fluctuations of the magnetic field at zero or double quantum frequencies, and therefore the NOE depend on the rate at which molecules reorient in solution. Only reorientation or "tumbling" is important, vibrations are too rapid, and translational motion involves large inter-nuclear distances. The dominance of $W_{2IS}$ or $W_{0IS}$ relaxation mechanism depends on $\tau_c$ of the molecule. $W_{2IS}$ dominates in small molecules with short $\tau_c$ and $W_{0IS}$ dominates in macromolecules

with long $\tau_c$ (proteins). This is due to the fact that the fluctuating magnetic fields that induce such relaxation processes need to contain frequencies that are equal to the sum of $\omega_I$ and $\omega_S$ (double quantum transitions), or to their small difference (zero quantum transitions). The size of the NOE effect is therefore also dependant on the magnetic field, i. e. it will be different at 300 and 600 MHz.

The product $\omega\ \tau_c$ reflects the NOE properties:

-When $\omega\ \tau_c << 1$ ($W_{2IS} > W_{0IS}$), condition known as the "extreme narrowing limit", the maximum NOE is $\eta = 1/2\ \gamma_S/\gamma_I$ . This means that the molecule tumbles fast and we have positive enhancements.

-When $\omega\ \tau_c >> 1$ ($W_{2IS} < W_{0IS}$),, the maximum NOE is $\eta = -\gamma_S/\gamma_I$. This means that the molecule tumbles slowly and we have a negative enhancement (proteins).

-When $\omega\ \tau_c \sim 1$ ($W_{2IS} \sim W_{0IS}$), the NOE goes to zero. This happens for certain medium sized molecules. The NOE can also be defined in terms of the resulting intensity I{S}, of spin I when spin S is saturated and its normal intensity $I_0$:

$$\eta = (I\{S\} - I_0)/I_0$$

A NOE enhancement can be accomplished by two means :

- A saturation of a resonance that causes the corresponding $\alpha$ and $\beta$ states to be equally populated. In the steady-state NOE experiment [111, 119], one spin is selectively saturated by a long, low-power continuous wave irradiation prior to recording a spectrum. A reference spectrum without irradiation is also recorded and when subtracted from the spectrum, only resonances having NOEs with the saturated spin are observed.
- An inversion pulse (180°) that inverts the relative populations of $\alpha$ and $\beta$ states. In this so-called transient NOE experiment [111, 119], a resonance is selectively inverted with a selective 180° pulse. After a delay for NOE build up, a 90° pulse is applied and a one-dimensional spectrum is acquired. Again spectra acquired in experiments with and without the selective 180° pulse are subtracted.

When analysing NOESY spectra, one has to be aware of an indirect NOE transfer, also called spin-diffusion. This indirect effect is a three-spin effect; magnetisation is further transferred from spin two to spin three, despite the fact that there is no direct NOE between spin one and three. Spin diffusion is very effective in macromolecules and steady state NOE measurements are therefore not used.

The transient NOE enhancements are symmetrical. In transient NOE experiments on large molecules, direct NOE peaks can be recorded at short mixing times, and spin diffusion can be recognised from the build up curves of NOEs from NOESY experiments that are recorded with increased mixing times.

### 2.1.9 Referencing

The shielding constant is an inconvenient measure of the chemical shift, in practice we use a relative scale, and refer all signals in the spectrum to the signal of a particular compound:

$$\delta = \frac{(\upsilon - \upsilon ref)}{\upsilon ref} \text{ X } 10^6 \text{ ppm} \qquad (1)$$

In biomolecular NMR, referencing [120] is often done on the water signal and is dependent on pH, salt concentration, and temperature:

$$\delta_{H2O} = 4.776 - 0.0119\,(T-25) - 0.002\,(pH-7) - 0.09\,(\text{salt concentration}) \qquad (2)$$

where T is temperature in celsius, pH is the pH of the sample and the salt concentration is the molar concentration of salt in the sample.

From here, the absolute frequency at 0 ppm for $^1H$ can be calculated from the frequency of water using equation (1). This absolute frequency of $^1H$, $\nu_0^H$ is then used to calculate the absolute frequency of the other nuclei $^{13}C$ ( $\nu_0^C = \gamma_C/\gamma_H \; \nu_0^H$ ) and $^{15}N$ ( $\nu_0^N = \gamma_N/\gamma_H \; \nu_0^H$ ). Again, equation (1) can be used to calculated the chemical shift of reference of the heteronuclei with the help of the absolute frequencies previously calculated and the nuclear frequencies at the position of the carrier.

## 2.2 Multidimensional NMR

### 2.2.1 Models used for the description of NMR experiments

There are several approaches for the analysis of pulsed NMR experiments, such as the vector model, density matrix evolution, and product spin operators. The vector model satisfactorily describes many basic NMR experiments but fails for example in the description of experiments involving multiple-quantum coherences. The most general description of a spin system is via evolution of a density matrix, which can handle also high order spin effects.

The initial state of a spin system is described by the spin density matrix $\rho(t)$ [121]. If we discard relaxation processes, the time evolution of the spin density matrix is described by the Liouville-Von Neumann equation:

$$\frac{d\rho}{dt} = -\frac{i}{h}[H, \rho]$$

Where H is a Hamiltonian that describes the different events happening in the system: chemical shift evolution, coupling and interaction with pulses. The spin density matrix describes the state of a large number of spins, and all observable and non-observable coherences can be extracted from the spin density matrix. The solution to the Liouville-Von Neumann equation will give the state of the magnetisation:

$$\rho(t) = e^{-iHt}\,\rho_0\,e^{iHt}$$

The calculation of the time dependent evolution of the system corresponds to transformation or rotation (can be expressed as a matrix) of the density matrix with the proper Hamiltonian-a pulse, chemical shift evolution, J coupling etc. This calculation is rather complicated, and hides the physical meaning of individual operations. It is more convenient to follow the evolution of a spin system on the basis of so-called product spin operators, which can, to a certain extent, be identified with the simple vector model.

### 2.2.2 General rules and description of product operators

When using some properties of matrix calculation, $\rho(t)$ becomes greatly simplified. If we call $\rho_0$, B, the matrix that describes the initial position of the magnetisation and H, A, a rotation (also a matrix) of angle $\omega t$, then we have $e^{-i\omega tA}$ B $e^{i\omega tA}$, and the following rules applied:

– A and B commute: $e^{-i\omega tA}$ B $e^{i\omega tA}$=B then thus A has no effect on B

– A and B anti-commute $e^{-i\omega tA}$ B $e^{i\omega tA}$= cos $\omega t$ B + sin $\omega t$ C where C=[A,B].

This means that B (initial magnetisation) has been rotated in respect to A by an angle $\omega t$ towards a new element C (Figure 2.12). The same description applies for a 90° pulse on the x axis where B corresponds to Iz, A to Ix and C is Iy=[Iz,Ix].

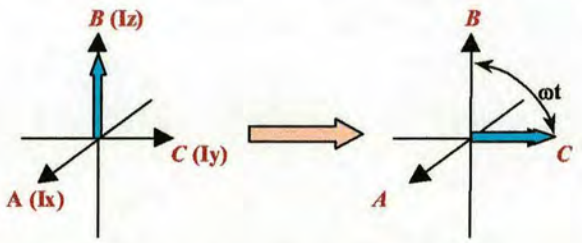
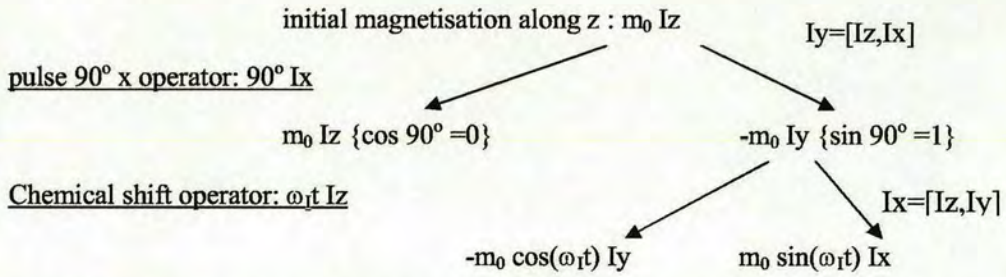



**Figure 2.12: A and B anti-commute**

Note: Pulse flip convention will be anticlockwise from now on.

Therefore for a simple 90°x pulse and acquisition, we have:




The effect of one of the evolution operators acting on one of the operators describing the state of the spin system where two ½ spins, I and S, are coupled, is summarise in Figure 2.13.

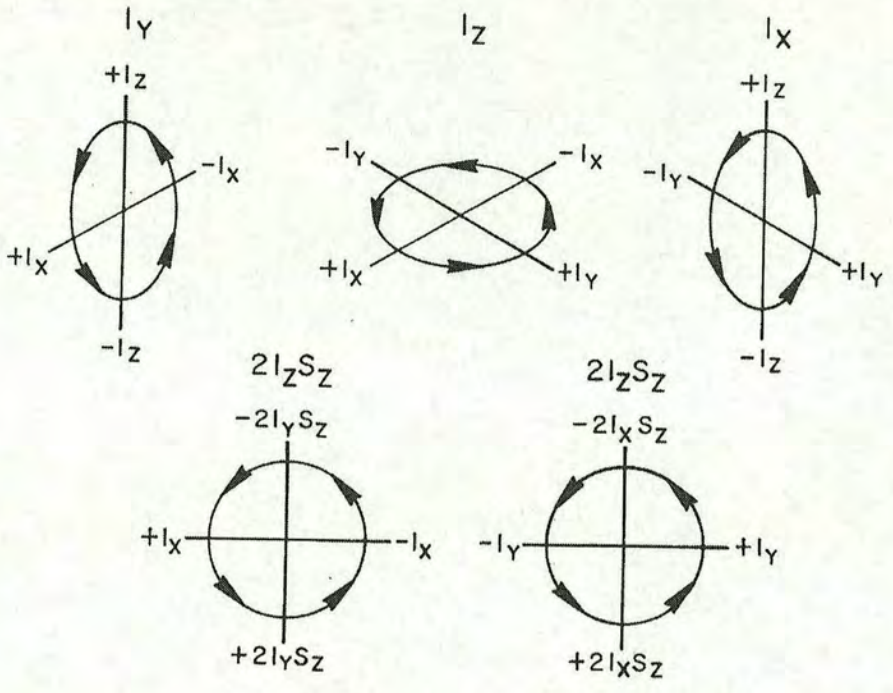



**Figure 2.13: Sign conventions for the evolution operators Iy, Iz, Ix and 2IzSz acting on the product operators Iy, Iz, Ix, 2IxSz and 2IySz.** ***Taken from reference*** **[112]**

Many building blocks of pulse sequences can be analysed using product spin operator. Effective Hamiltonians describing cumulative effects of different events lead to simple equations (Figure 2.14).

### 2.2.3 Quadrature detection

If detection is only carried out along one axis in the rotating frame then it is impossible to discriminate between the signs of frequencies i.e. determine the sense of precession of the magnetisation vectors. Use of two orthogonal detectors along the x and y-axes allows detection of the sense of rotation. This is known as quadrature detection [122]. The signals are sampled simultaneously by two analogue-to-digital converters (ADC), or if we cannot digitise the two FIDs simultaneously, then only one of each FID is digitised at a time, alternately. These two FIDs are then used as the real and imaginary parts for complex FT.

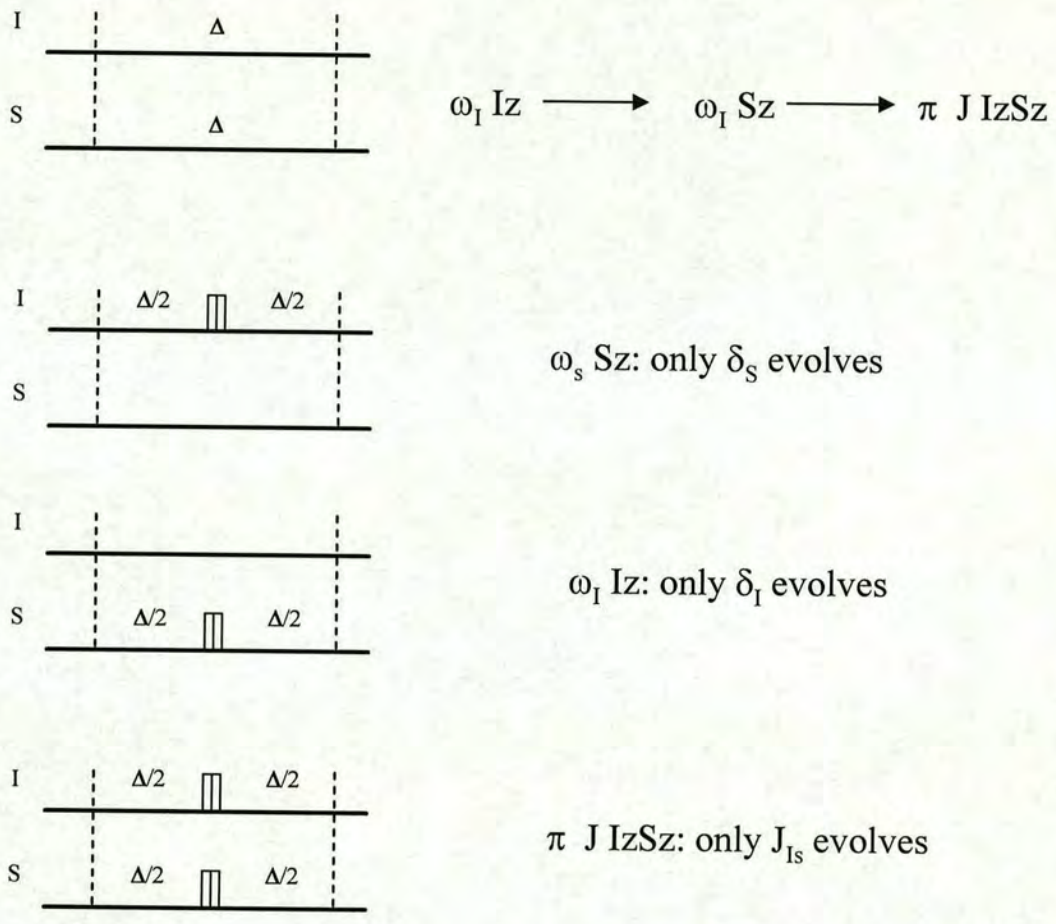



**Figure 2.14: Description of the reduced product operators (double bars represent 180° pulse)**

In practice the placement of the detectors in the rotating frame is dependant on the phase of the carrier frequency which is subtracted from the signal: two detectors in quadrature are fed with the same absolute frequency, but one is shifted in phase by 90°. The acquisition of real and imaginary data points can be carried out either in a simultaneous or alternative fashion. In the alternative mode, also known as the Redfield method, real and imaginary points are sampled alternately by switching the detector phase by +/- 90°; this is equivalent of moving the detector around the axes of the rotating frame. Compared to the simultaneous manner, or true complex acquisition, the alternative mode has a dwell time half as long (1/2 SW). When using the alternative mode, after FT, the spectrum has twice the SW of the spectrum that would be obtained with true complex acquisition, it also contains mirror images of all signals which are symmetric to SW/2. This effectively places the carrier frequency at the

edge of the spectrum when compared to true complex acquisition mode. We obtain two mirror images of the spectrum but we just discard the half outside the chosen spectral range.

The methods described above allow sign discrimination in the acquired dimension, but in multidimensional spectroscopy we also need sign discrimination in the indirectly acquired dimension. Again we have two methods. The States (-Ruben-Haberkorn) method is analogous to the true complex acquistion mode in 1D. For each time increment in the indirectly acquired dimension, the pulse sequence is performed twice with the phases of the pulses adjusted such that we observe the magnetisation from two orthogonal viewpoints. The phase of the first pulse is shifted by 90°. The second method is analogous to the Redfield method and is known as the TPPI (Time Proportional Phase Incrementation) method. Only the real data points are recorded in the indirectly detected dimension, but twice as many, with half the time increment.

### 2.2.4 Composite pulses and broadband decoupling

High resolution NMR requires perfect pulses. In order to compensate for pulse imperfection e.g. radiofrequency inhomogeneity, resonance offset error (precession of the magnetisation around an effective $B_1$ field), and miscalibration; composite pulses are employed. A composite pulse is a combination of different pulses that achieve the same effect that would be obtained from a single pulse, but here the imperfections are compensated for and we have a better result [123]. Composite 180° pulses are routinely used in modern NMR to achieved basic tasks like spin-locking, and broadband decoupling. One of the first composite 180° pulses to be developed was the following pulse scheme: 90° x -180° y - 90° x, and was further used as a buliding block in more complex composite pulses. A common composite pulse is the MLEV-16 sequence [124], where several repeats of this block and its phase inverted equivalent are put together. This sequence allows isotropic mixing and is used in TOCSY experiments. Composite 180° pulses are also used for broadband decoupling, particularly in heteronuclear experiments.

### 2.2.5 Pulsed field gradients

The use of pulsed magnetic field gradients (PFG) is nowadays almost compulsory in high resolution NMR. Fields gradients are used to destroy temporarily the homogeneity of the magnetic field in a controlled manner and perform a variety of functions in pulse sequences. A PFG consists of an additional pair of coils inside the probe, on both sides of the sample. When a direct current is sent through this coil pair in opposite directions, we have a resulting magnetic field parallel to $B_0$ on one side of the sample and antiparallel on the other side which create an overall linear field along the direction of the coil pair, i..e. along z if one coil is above and one below the sample (z gradient). Routine use of PFGs is due to the implementation of so called "shielded gradients" which increase the gradient strength without the need for long recovery delays and reduce the induction of eddy currents elsewhere in the probe. Typical gradient strength is of the order of 30 to 70 Gauss per cm. Their principal applications are the selection or destruction of coherences.

During the application of a PFG, how the resonance of a nucleus will be affected depends on its spatial location. The phase twist $\Delta\varphi_G$ caused by the gradient is given by $\Delta\varphi_G = \Delta\omega_G \tau_G = \gamma B_G \tau_G$, [113] where $\tau_G$ is the gradient duration, $\Delta\omega_G$ is the change in precession frequency caused by the gradient, and $B_G$ is the gradient strength. A gradient applied on a transverse magnetisation results in twisting its phase at the edges of the sample several hundred times and the detectable net magnetisation is null. The signal defocused by a PFG cannot completely be refocused by a second gradient if it has diffused away between the two gradients. This has implications for gradient spin echoes where a gradient is performed before and after a 180° pulse. Since the 180° pulse reverses the twist caused by the first gradient, a second gradient of equal strength and sign is used to refocus the signal. The result is a clean up of the magnetisation components that were not refocused by the 180° pulse due to miscalibration. Such spin echoes are used in diffusion experiments where the two gradients are separated by a variable delay to allow measurement of the molecular diffusion, and to a certain extent in several water suppression schemes.

PFGs are also useful to destroy unwanted magnetisation components in pulse sequences. For example, when using z gradients, only the x and y components of the magnetisation are affected and all z components remain unaffected. This is of particular interest in heteronuclear coherence transfer steps where the magnetisation of interest is in a zz-ordered state during the application of the gradient.

The third category of applications, coherence selection, requires precisely reproducible gradients. The phase twist caused by a gradient depends on its length and strength but also on the type of coherence it affects. $^1H$ coherences dephase four times as fast as $^{13}C$ coherences under the

same gradient and $^1H$-$^1H$ double-quantum coherences dephase twice as fast as $^1H$ single-quantum coherences under the same gradient. These features are used to selectively refocus only specific combinations of coherences with a pair of gradients.

### 2.2.6 Water suppression

Water signal suppression is of particular interest in biomolecular NMR. Signals of interest arise from protein which usually has a concentration of half to few mM, while the water signal arises from protons at a concentration of 110 M. Thus, several methods have been developed to reduce the signal produced by water: presaturation, semi-selective excitation, trim pulses and the use of PFGs. Best solvent suppression is generally achieved with PFG techniques.

#### 2.2.6.1 Prestaturation

The most common method of water suppression is selective presaturation. In this method the populations of the two energy states for water protons are equalised, resulting in no net magnetisation. This is accomplished by a rf pulse set on the water resonance frequency for a long period of time (1-2 s), and effectively in a vector model causes the solvent magnetisation to rotate in the vertical plane. Because of the radio frequency inhomogeneity of the presaturation pulse, the magnetisation is randomly distributed in this plane by the end of the pulse with no net component in the xy plane. Presaturation is performed immediately prior to the pulse sequence. Presaturation is not the best water suppression technique since it often saturates protein frequencies close to water, or protein frequencies in exchange with water like NH protons. The SCUBA (Stimulated Cross-relaxation Under Bleached Alphas) [125] sequence is of particular use to avoid such problems.

#### 2.2.6.2 Semi-selective excitation

A better solvent suppression method than using presaturation can be achieved with semi-selective excitation. If we place the carrier frequency on water, then after a 90° x pulse, water will be static along the y axis in the rotating frame, while protein resonances will be precessing. After a delay, a second 90° -x pulse will then return the water magnetisation along z, and the protein resonances will be in the xz plane. Their x component that evolved during the delay remains and therefore can be acquired without any contribution from water. Such scheme is known as the jump-and-return sequence [126] and is performed at the end of pulse sequence, prior to acquisition.

### 2.2.6.3 Trim pulses

Trim pulses use for solvent suppression are of a few milliseconds duration and applied along the x or y axis, effectively spin locking the magnetisation along an axis while rf inhomogeneity dephases the magnetisation not aligned with the spin lock. Trim pulses are also useful to dephase unwanted coherence in pulse sequences.

### 2.2.6.4 Use of pulsed field gradients

A popular alternative for solvent suppression is the use of PFG. Several methods using PFG have now been developed. In the Watergate scheme [127], a combination of water-selective and non-selective pulses with PFG is used at the end of the pulse sequence. This building block ensures that water is completely dephased while the protein signal is refocused. Another, very effective water suppression scheme based on PFG uses the DPFGSE (Double Pulsed Field Gradient Spin Echo) sequence [128] which has been used in some experiments performed for this study.

## 2.2.7 Phase cycling

Phase cycling involves shifting the phase of the pulses and/or the receiver by a controlled amount between individual scans. This is used to achieve coherence selection, and to cancel out residual magnetisation arising from pulse imperfections, as well as, to perform quadrature detection in the indirectly acquired dimension.

A system has coherence when its elements, here the spins, have a common relationship, in this a common phase. At equilibrium, the magnetisation is not coherent since every spin is precessing with a random phase. In contrast after a 90° pulse, a system of spins is coherent, and $T_2$ relaxation can be explained as a loss of coherence. In a one-spin system, we have two possible states, $\alpha$ and $\beta$, and consequently a change of only one quanta is involved; this is a single-quantum transition, and the coherence order (p) equals +/- 1. Single quantum coherences are the only type of coherence that generates a current in the receiver coil, and they are also known as observable coherences. We can select to detect p = -1 by selecting the phase of the pulses and receivers. A 90° pulse excites single-quantum transitions and its effect can be represented in a diagram also known as a coherence transfer pathway. By convention, coherence order $p = -1$ is associated with $+\omega_0$ and $p = +1$ with $-\omega_0$. In a two-spin system, we have four possible states, and have zero-, single-, and double-quantum

transitions; p = 0, p = +/-1, and p = +/-2 respectively. It is also possible to generate higher order coherences which can be selected using Bodenhausen's rules [129].

### 2.2.8 Two-dimensional NMR experiments

NMR experiments can be subdivided into four stages: preparation, evolution, mixing, and detection.

#### 2.2.8.1 Homonuclear two-dimensional correlated spectroscopy or COSY experiment

The COSY experiment (Figure 2.15) [130] is based on the transfer of coherence from one spin to another via J coupling.

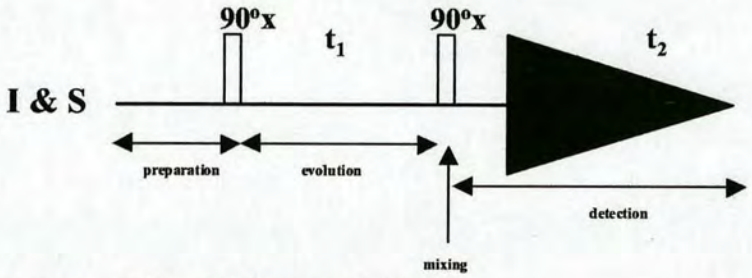
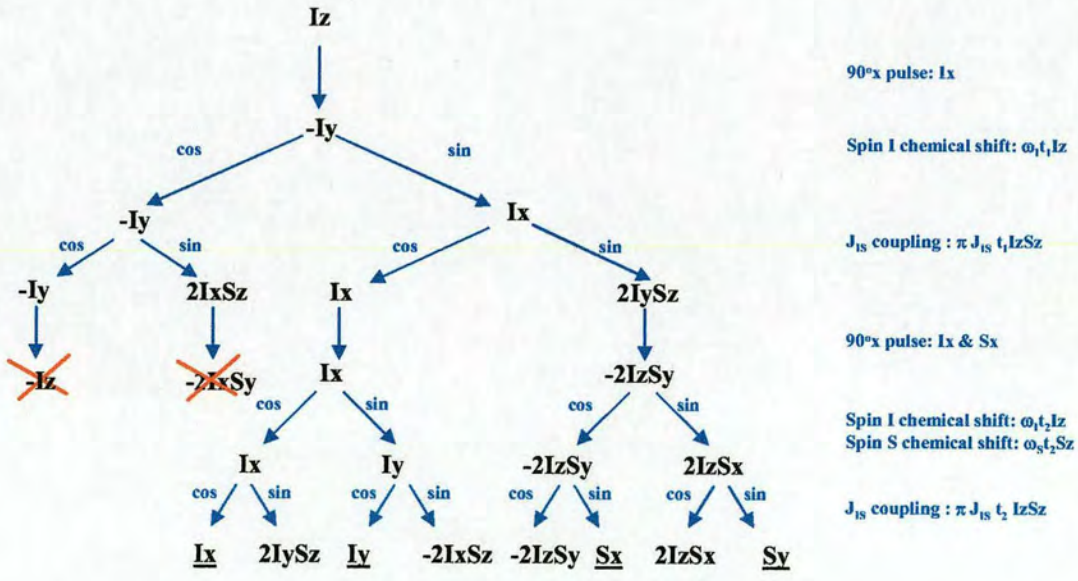



**Figure 2.15: Pulse sequence of the COSY experiment**

If we consider spin I:



Only the underlined coherences are detectable, thus if we detect the signal on the y axis, we will have the following signals coming from spin I:

| | | | | |
|---|---|---|---|---|
| Iy modulated by: | $\cos \pi J_{IS} t_1$ | $\sin \omega_I t_1$ | $\cos \pi J_{IS} t_2$ | $\sin \omega_I t_2$ |
| Sy modulated by: | $\sin \pi J_{IS} t_1$ | $\sin \omega_I t_1$ | $\sin \pi J_{IS} t_2$ | $\sin \omega_s t_2$ |

Signals coming from spin S generated from Sz in the same way:

| | | | | |
|---|---|---|---|---|
| Sy modulated by: | $\cos \pi J_{IS} t_1$ | $\sin \omega_s t_1$ | $\cos \pi J_{IS} t_2$ | $\sin \omega_s t_2$ |
| Iy modulated by: | $\sin \pi J_{IS} t_1$ | $\sin \omega_s t_1$ | $\sin \pi J_{IS} t_2$ | $\sin \omega_I t_2$ |

The Iy and Sy signals coming from spin I, and S, respectively, will give rise to the diagonal peaks. In each dimension, we have the product of a cosine modulation (J modulation) and a sine modulation (precession), which after FT corresponds to their convolution product. The real part of this product has the shape of inphase doublets in dispersion (d).

The Iy and Sy signal coming from spin S, and I, respectively, will give rise to the cross-peaks of the COSY spectrum. This time we have for each dimension the product of two sine modulations. The real part of their convolution product has the shape of antiphase doublets in absorption (a). The final result is a two-dimensional spectrum as shown in Figure 2.16.

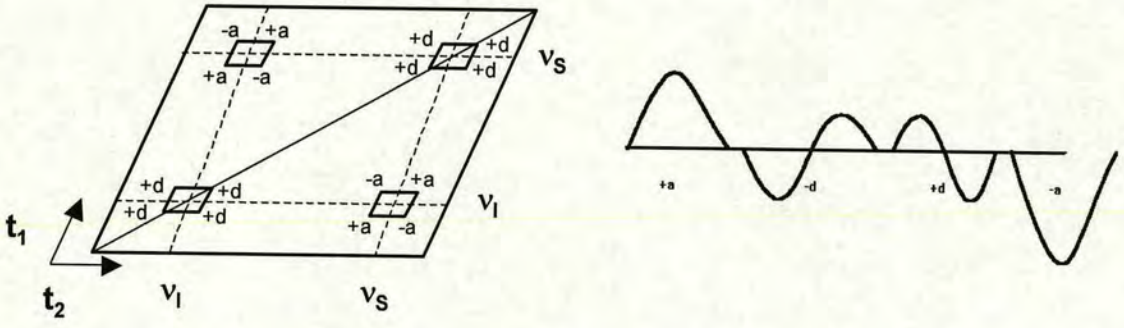



**Figure 2.16: Schematic COSY 2D spectrum and its associated signal lineshapes [131]**

The double quantum filtered COSY sequence uses an extra 90° pulse at the end of the COSY sequence and an appropriate phase cycling to select double quantum coherence after the 90° second pulse. In this sequence, the diagonal peaks will be antiphase multiplets with absorption lineshape and this will avoid the long tails on peaks seen in the COSY experiment that make difficult to observe the crosspeaks near the diagonal.

### 2.2.8.2 Total correlation spectroscopy (TOCSY)

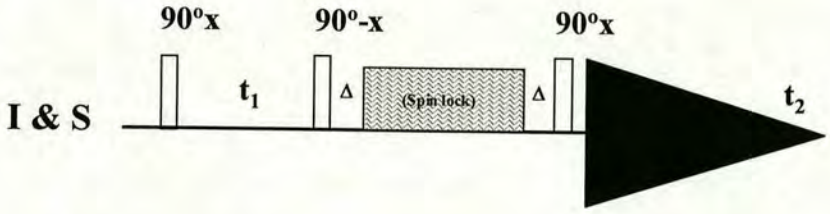



**Figure 2.17: Pulse sequence of the TOCSY experiment.**

The TOCSY experiment [132] is also known as homonuclear Hartmann Hahn experiment (HOHAHA) [133] . The TOCSY experiment is used in the $^1$H-NMR of proteins NMR in conjunction with COSY experiment for identification of resonances that are all within one residue or spin-system. The key feature of the TOSCY experiment is that it uses a period of spin locking to achieve coherence transfer (Figure 2.17). A spin lock pulse is a radio frequency pulse with a duration up to tens of milliseconds; more precisely it is a train of non-selective 180° composite pulses. This means that during the spin lock, the only active interactions are the coupling constants, and there will be a transfer of magnetisation between two coupled spins when the Hartmann-Hahn condition is satisfied. This condition implies that the difference in the magnetic field that they experience is less than the magnitude of their J coupling. Usually, if we consider two spins coupled by a spin-spin interaction $J_{IS}$, after creating transverse magnetisation with a 90°x pulse, we have a weak-coupling Hamiltonian, $2\pi J_{IS}$ IzSz; spin locking of both I and S means that they will evolve under a strong coupling term, $2\pi J_{IS}$ (IxSx+IySy+IzSz).

The evolution of a coherence Ix, will be as follows:

$$Ix \xrightarrow[\text{Strong}]{} Ix\,(1+\cos 2\pi J_{IS}\tau)/2 + Sx\,(1-\cos 2\pi J_{IS}\tau)/2 + (IySz-IzSy)\sin 2\pi J_{IS}\tau$$

where $\tau$ is the duration of the mixing time.

There are two different classes of spin-locks: isotropic and anisotropic. Isotropic spin-locks (WALTZ-16 or MLEV-16) allow transfer of all magnetisation components, while anisotropic spinlocks (MLEV-17) can transfer only one transverse component and the z component. Only anisotropic spinlock sequences are employed for TOCSY experiments; the initial anisotropic version of the TOCSY experiment was named the homonuclear Hartmann-Hahn (HOHAHA) experiment, and today the terms TOCSY and HOHAHA are mostly used interchangeably.

At the end of the $t_1$ period of the TOCSY pulse sequence in Figure2.19, we have:

$-Iy \cos \omega_I t_1 \cos \pi J_{IS} t_1 + 2IxSz \cos \omega_I t_1 \sin \pi J_{IS} t_1 + Ix \sin \omega_I t_1 \cos \pi J_{IS} t_1 + 2IySz \sin \omega_I t_1 \sin \pi J_{IS} t_1$.

Before entering the spinlock, we can select the y component by using appropriate phase cycling. Only $-Iy \cos \omega_I t_1 \cos \pi J_{IS} t_1$ and $2IySz \sin \omega_I t_1 \sin \pi J_{IS} t_1$ terms are considered here. Focusing on the first term, a 90° –x pulse flips this –Iy component back onto the z axis.

During spin-lock, mixing occurs and the z component gives rise to:

$Iz \cos \omega_I t_1 \cos \pi J_{IS} t_1 (1+\cos 2\pi J_{IS}\tau)/2 + Sz \cos \omega_I t_1 \cos \pi J_{IS} t_1 (1-\cos 2\pi J_{IS}\tau)/2 + (IxSy-IySx) \cos \omega_I t_1 \cos \pi J_{IS} t_1 \sin 2\pi J_{IS}\tau$, where $\tau$ is the duration of the mixing time.

Finally, the Iz and Sz terms are converted into transverse, observable magnetisation with a 90° x pulse, and give rise to diagonal and cross peaks respectively. The multiple quantum terms created during the spin-lock are converted into antiphase magnetisation, which during the $t_2$ period evolves into observable magnetisation. In the spectra, this gives rise to antiphase multiplets, which in protein spectra are strongly diminished by mutual cancellation of their broad lines with opposite phase.

The mixing period induces an undesirable transfer via dipolar coupling, also called, rotating Overhauser effect (ROE). ROE peaks are of opposite sign to those of nuclear Overhauser peaks, thus if we allow NOE peaks to build up when the magnetisation is aligned along the z axis before and after the spin-lock, NOE peaks will then compensate for ROE peaks. This is achieved by setting the delays $\Delta$ to be equal to $\tau/4$ as the build-up rate of ROE peaks is twice that of NOE peaks. Alternatively one can use "clean" isotropic mixing sequences (DIPSI-2rc) [134] that include directly the delays within the composite pulses, when the magnetisation is aligned along the z axis to permit compensatory NOE cross-relaxation.

### 2.2.8.3 Nuclear Overhauser enhancement spectroscopy (NOESY)

Like the COSY or TOSCY experiments, the NOESY experiment [135] starts with a 90° x pulse followed by an indirect evolution time (Figure 2.18), at the end of which the following terms exist:

$-Iy \cos \omega_I t_1 \cos \pi J_{IS} t_1 + 2IxSz \cos \omega_I t_1 \sin \pi J_{IS} t_1 + Ix \sin \omega_I t_1 \cos 2\pi J_{IS} t_1 + 2IySz \sin \omega_I t_1 \sin \pi J_{IS} t_1$.

A second 90° x pulse leads to:

$-Iz \cos \omega_I t_1 \cos \pi J_{IS} t_1 - 2IxSy \cos \omega_I t_1 \sin \pi J_{IS} t_1 + Ix \sin \omega_I t_1 \cos \pi J_{IS} t_1 - 2IzSy \sin \omega_I t_1 \sin \pi J_{IS} t_1$.

The phase cycling is designed to select only coherence order p=0 during the mixing time i.e. the first term (-Iz) and the zero quantum part of the second term (1/2[IySx-IxSy]).

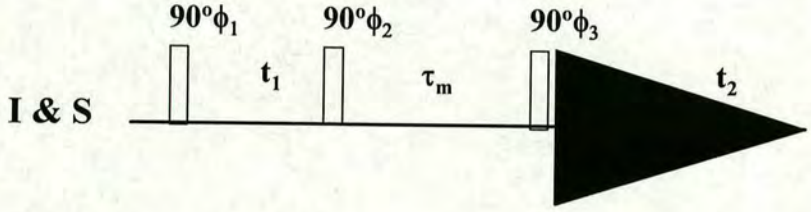



**Figure 2.18: Pulse sequence of the NOESY experiment**

This first term is a z magnetisation, a non-equilibrium polarisation, modulated with the cosine of the chemical shift and coupling, which causes an NOE on neighbouring nuclei:

$$-Iz \cos \omega_I t_1 \cos \pi J_{IS} t_1 \xrightarrow[\text{Mixing}(\tau_m)]{} -Iz(1-R_{II}\tau_m) \cos \omega_I t_1 \cos \pi J_{IS} t_1 - Sz(R_{IS}\tau_m) \cos \omega_I t_1 \cos \pi J_{IS} t_1$$

where $R_{IS}$ is the cross-relaxation rate between spin I and S, $R_{II}$ is the auto-relaxation rate of spin I, and $\tau_m$ is the mixing time.

Cross relaxation rate and the duration of $\tau_m$ determine the amount of NOE transfer. After the final 90° x pulse, these z components are converted into detectable signals: Iy $(1-R_{II}\tau_m)$ cos $\omega_I$ $t_1$ cos $\pi J_{IS}$ $t_1$ + Sy $(R_{IS}\tau_m)$ cos $\omega_I$ $t_1$ cos $\pi J_{IS}$ $t_1$. The first term results in a diagonal peak after 2D FT (at $\omega_I$), the second one in a cross-peak at $\omega_I$ in F1 ($t_1$) and at $\omega_S$ in F2 ($t_2$). The sign of the cross-peak depends on the sign of the cross relaxation rate, i.e. the dominance of either $W_{0IS}$ or $W_{2IS}$.

The second term mentioned above results in the unwanted zero-quantum diagonal and cross peaks that are dispersive antiphase signals in both dimension, and some unobservable signal. These peaks can be suppressed by recording and co-adding several NOESY spectra with a range of values of $\tau_m$. In large molecules, they can often be ignored as the zero quantum relaxation rate is very fast, the antiphase contributions tend to cancel, and they don't affect the integrated volume of the cross-peaks (antiphases don't have a net integral).

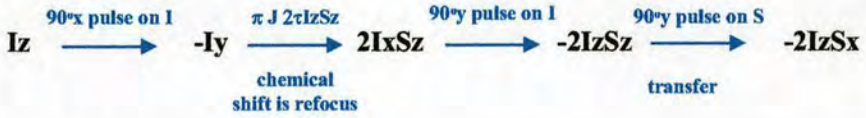
The initial slope of the NOE as a function of $\tau_m$ (build-up curve) corresponds to the cross relaxation rate, which depends on the internuclear distance. This can be measured in 2D NOESY with a sufficiently short $\tau_m$, where NOE cross-peak integrals can be converted into internuclear distances. This is usually done with a reference NOE corresponding to a known distance: $r_{ij} = r_{ref} (I_{ref}-I_{ij})^{-6}$, where $r_{ij}$ is the distance between two atoms, i and j, and $I_{ij}$ the integral of their NOE cross-peak.

### 2.2.8.4 Heteronuclear single quantum spectroscopy (HSQC)

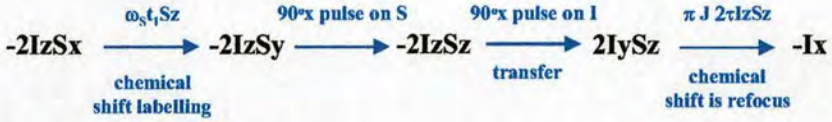
Heteronuclear correlation spectroscopy is based on the same principles outlined in the COSY experiment, but with some important differences. In heteronuclear correlation experiments, the coherences from $^1H$ are frequency-labelled during $t_1$ by the heteronucleus ($^{15}N$,$^{13}C$ etc), and both frequency dimensions are decoupled to concentrate intensity on singlets at the chemical shift frequencies. The sensitivity of this type of experiments is dependent on the gyromagnetic ratio as: Signal/Noise ~ $\gamma_I$ $(\gamma_s)^{3/2}$ [113], for two spins, I and S. Thus this equation shows that starting with the nucleus with the highest $\gamma$ but also detecting that same nucleus achieves the highest sensitivity. There are two principle schemes for transferring magnetisation between coupled heteronuclei resulting in the creation of single and multiple-quantum coherence transfer, named heteronuclear single-quantum spectroscopy (HSQC) [136] and heteronuclear multiple-quantum spectroscopy (HMQC) [137] respectively. HMQC requires fewer 180° pulses and therefore is less sensitive to pulse miscalibration and radio-frequency inhomogeneity. Nevertheless single-quantum transfer is generally used because

of the favorable relaxation properties of single quantum coherences and singlet structures in the spectra.

The HSQC experiment employs two INEPT (Insensitive Nucleus Enhancement by Polarisation Transfer) building blocks (Figure 2.19). The first INEPT is a preparation stage to achieve antiphase heteronuclear coherence:




The second INEPT block is described as a reverse INEPT because it converts this coherence back to $^1$H magnetisation after labelling under the heteronuclear chemical shift:




The delays $\tau$, and $\Delta$ are set to 1/4 $J_{IS}$, and 1/8 $J_{IS}$ respectively so that the transfer of magnetisation from I to S is maximised. A broadband decoupled spectrum is acquired so heteronuclear coupling is again supressed, and $^1$H chemical shift will develop. To select only heteronuclear-bound protons, a phase cycling scheme is used on the spin S 90° pulse in the reverse INEPT block. Protons that are not directly bound to the heteronucleus will not develop into –2IzSy terms and therefore not be affected by phase changes of the spin S 90° pulse.

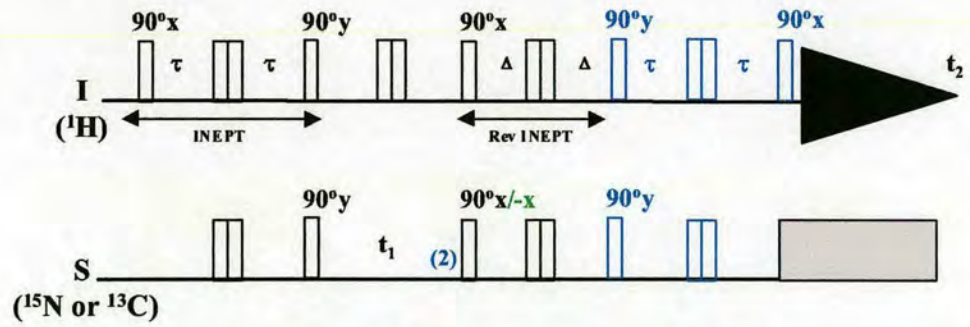
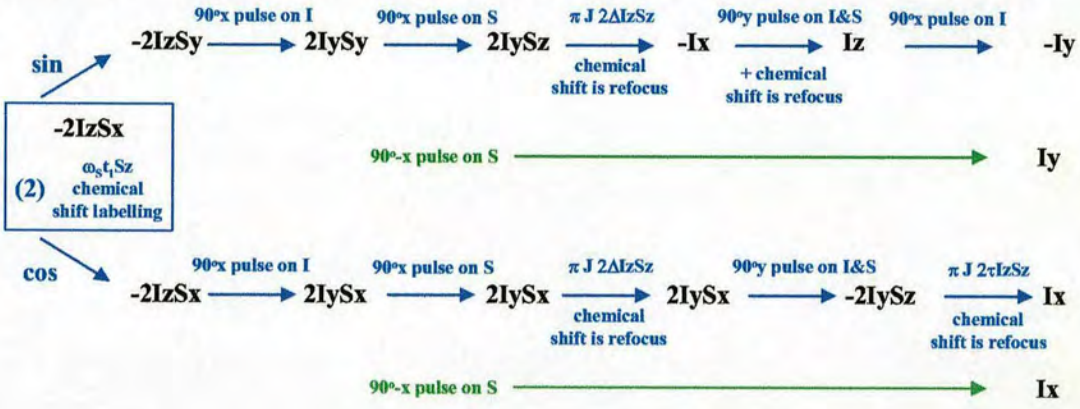



**Figure 2.19: Pulse sequence of the SE-HSQC experiment.**

By subtracting two scans acquired with such phase cycling, signal from non-heteronucleus bound protons will cancel. This experiment can be conduced in a sensitivity enhanced (SE) manner [138], which yields a phase modulated signal, but with a factor $(2)^{1/2}$ increase. The SE-HSQC starts like a regular HSQC experiment. We have the first INEPT block to create antiphase coherence, then labelling with spin S chemical shift during $t_1$:




When recording a SE-HSQC in the TPPI manner, one acquires cosine and sine modulated signal in $t_1$, to which both coherence transfer pathways have contributed and which has a factor of $(2)^{1/2}$ higher signal to noise ratio. Phase cycling of the second 90° heteronuclear pulse allows control of the sign of Iy. These two FIDs are combined to yield pure sine and cosine modulated signals, with twice the intensity, which then can be used as input for complex FT.

### 2.2.9 Three dimensional spectroscopy

The assignment strategy for unlabelled proteins uses a combination of COSY, TOCSY and NOESY experiments. However resonance overlap becomes so severe in two-dimensional spectra of proteins with more than 80 residues, that it is often impossible to identify adjacent spin system. In three-dimensional experiments, three frequencies rather than two label each crosspeak. When using $^{15}N$ labelling, it is possible to resolve the resonances of the amide protons with the $^{15}N$ chemical shift of their attached nitrogen. This is used in three dimensional TOCSY-($^{15}N$, $^{1}H$) HSQC [139] and NOESY-($^{15}N$, $^{1}H$) HSQC [139-141] experiments. Although this facilitates sequential assignment, the sequential connectivity between spin-systems still relies on identification of the Hα of each residue and therefore has some ambiguities. The limited resolution among Hα resonances is overcome in triple resonance experiments used in $^{15}N/^{13}C$ labelled proteins as they employ coherence transfer through the $^{1}J$ or $^{2}J$ heteronuclear coupling (large coupling) only rather than the $^{3}J$ homonuclear coupling (small coupling) used in COSY and TOCSY experiments [142].

In two-dimensional experiments, the evolution period labels spins with their chemical shifts in the indirectly detected dimension, while in the mixing period spins are correlated to each other. In three-dimensional experiments, the pulse sequence is constructed by combining two 2D experiments without the detection period after the first 2D block. One can think of this initial 2D sequence as a preparation period for the second 2D part. Thus two evolution periods are incremented independently to give a data matrix with three different acquisition times: $t_1$, $t_2$, $t_3$.

#### 2.2.9.1 3D $^{15}N$ edited TOCSY-HSQC and 3D $^{15}N$ edited NOESY-HSQC

$^{15}N$ edited TOCSY-HSQC and $^{15}N$ edited NOESY-HSQC are concatenations of a homonuclear TOCSY and NOESY, respectively, with a ($^{15}N$, $^{1}H$) HSQC (Figure 2.20). In these experiments only NOESY and TOCSY peaks of amide protons are detected. Here the amide protons are acquired in the directly detected dimension ($t_3$), while their NOEs are indirectly detected ($t_1$) to discard water signal contributions in the NOE peaks [142]. If the order of individual building blocks were reversed, i.e. in a 3D- ($^{15}N$, $^{1}H$) HSQC-NOESY, the water signal would obscure NOE peaks identification.

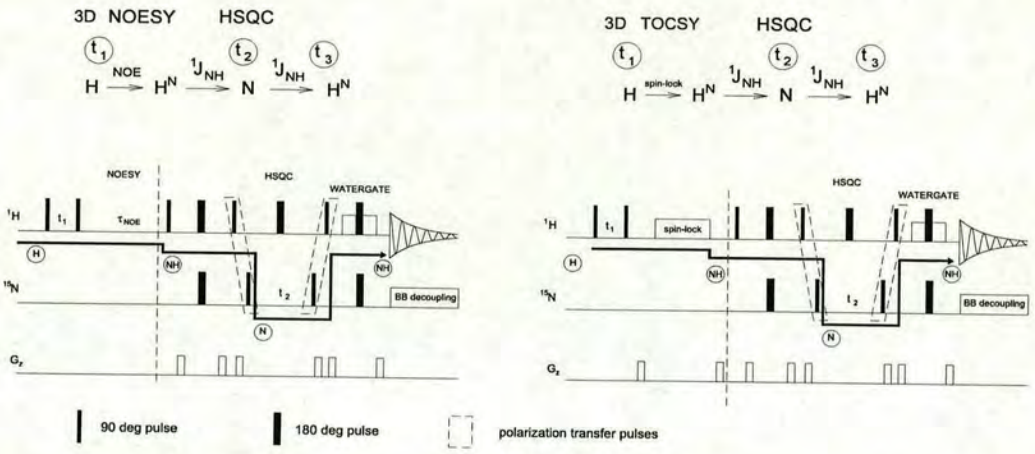



**Figure 2.20: the $^{15}N$ edited TOCSY-HSQC and $^{15}N$ edited NOESY-HSQC pulse sequences.** ***Taken from reference [142].***

### 2.2.9.2 Triple resonance experiments

The nomenclature of the triple-resonance experiments reflects the magnetisation transfer pathways of the experiments, and the nuclei that are involved form the name of the experiments [143]. Nuclei, whose chemical shifts are not labelled during the indirectly detected dimensions, but through which magnetisation passes, are put in parentheses. Triple-resonance experiments can be divided into two groups: experiments that are used for backbone assignment (HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, HBHA(CO)NH, HBHANH etc), and experiments that are used for side chain assignment (H(C)(CO)NH-TOCSY, (H)C(CO)NH-TOCSY). Backbone experiments are usually analysed in pairs (HNCO with HN(CA)CO, CBCA(CO)NH with HNCACB, and HBHA(CO)NH with HBHANH), to provide sequential assignment of spin systems. The assignments deduced from the analysis of the triple-resonance experiments are then used in double-resonance experiments ($^{15}N$ edited NOESY-HSQC, H(C)CH-TOCSY and $^{13}C$ edited NOESY-HSQC). For illustration purposes, one triple resonance experiment – the HNCO – is considered here in detail.

### 2.2.9.2.1 HNCO

The HNCO experiment (Figure 2.21) correlates the backbone amide proton and nitrogen frequencies of one residue with the carbonyl frequency of the preceding residue in the protein sequence.

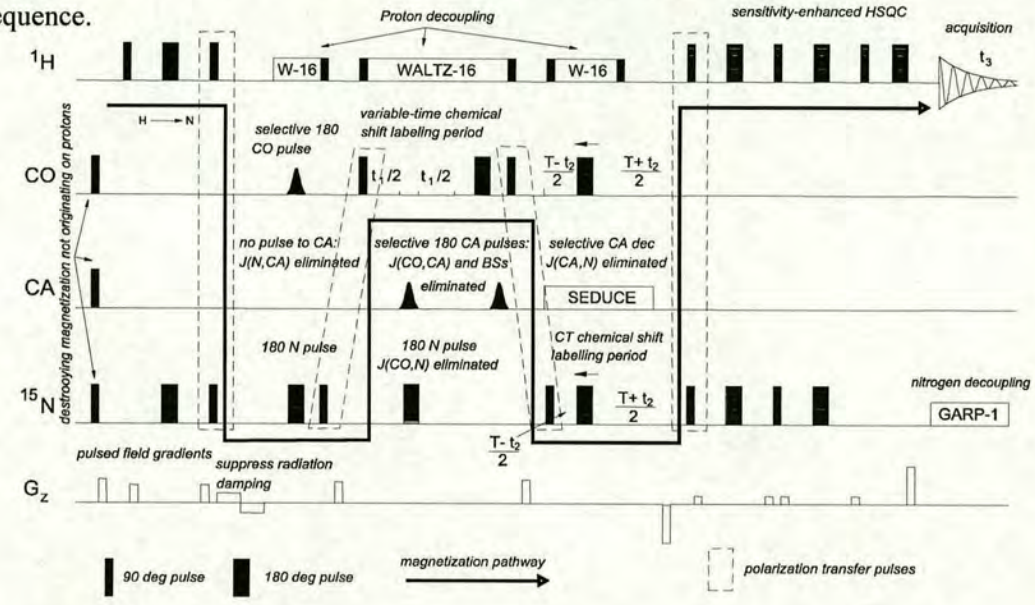



**Figure 2.21: the 3D HNCO pulse sequence.** ***Taken from reference*** **[142]**

Only these correlations are observed in the HNCO since $^1J_{CO(i-1),N(i)} = 15$ Hz, while $^2J_{CO(i),N(i)} \sim 0$ Hz. The magnetisation-transfer pathway is the following:

$$^1H \xrightarrow{J_{NH}} {}^{15}N \xrightarrow{J_{NCO}} {}^{13}CO\,(t_1) \xrightarrow{J_{NCO}} {}^{15}N\,(t_2) \xrightarrow{J_{NH}} {}^1H\,(t_3)$$

The sequence starts with an INEPT transfer of the magnetisation from the backbone $^1$H to $^{15}$N. Next, a delay during which only the one-bond coupling between $^{15}$N and $^{13}$CO is allowed, builds up antiphase two-spin coherence NxCOz. This is further converted into NzCOy by a selective $^{13}$CO 90° pulse and $^{15}$N 90° pulse for $^{13}$CO-chemical shift labelling during the evolution time $t_1$. A second pair of selective $^{13}$CO 90° pulse and $^{15}$N 90° pulse converts the magnetisation into NyCOz. This coherence then evolves (during the delay 2T) into $^{15}$N coherence in antiphase with respect to $^1$H; HzNy, which becomes simultaneously labelled by the $^{15}$N-chemical shift during the constant-time $^{15}$N evolution period $t_2$. Finally, a reverse INEPT transfer with sensitivity enhancement converts the coherences back to the amide protons for observation.

#### 2.2.9.2.2 Other protein backbone and side chain assignment experiments

The HN(CA)CO experiment correlates the chemical shifts of the backbone amide proton (i), nitrogen (i) and two carbonyl frequencies (i, i-1). Therefore when used along with the HNCO experiment, this pair of experiment allows sequential assignment of the spin systems in a $^{15}N/^{13}C$ – labelled protein. Similarly the pair of experiments, CBCA(CO)NH - HNCACB, and HBHA(CO)NH - HBHANH are used for sequential assignment. The H(C)(CO)NH-TOCSY, and (H)C(CO)NH-TOCSY experiments are further used to assigned the remainder of the proton and carbon side-chain resonances respectively.

A schematic description of the different frequencies observed in these 3D-experiments along with the way they are displayed in the Ansig software package for resonance assignment can be found in figures 2.22 and 2.23.

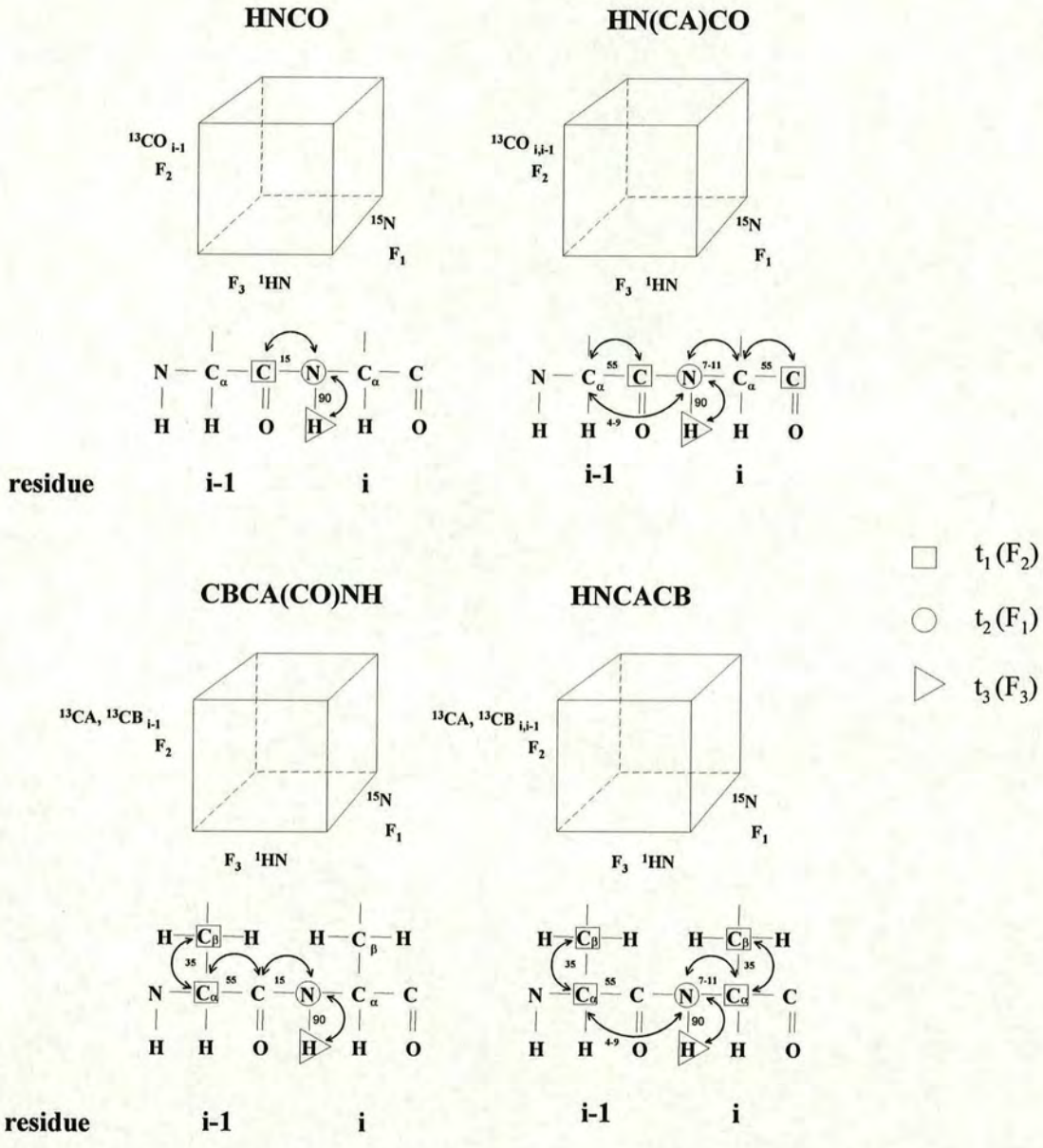



**Figure 2.22: Experiments for backbone assignments in 15N/13C-labelled proteins**

Experiments are described as they are viewed in the Ansig software package for resonance assignments. Note: the so-called "out and back" implementation was used where the magnetisation starts and finishes on NH protons, as indicated by the double-headed arrows.

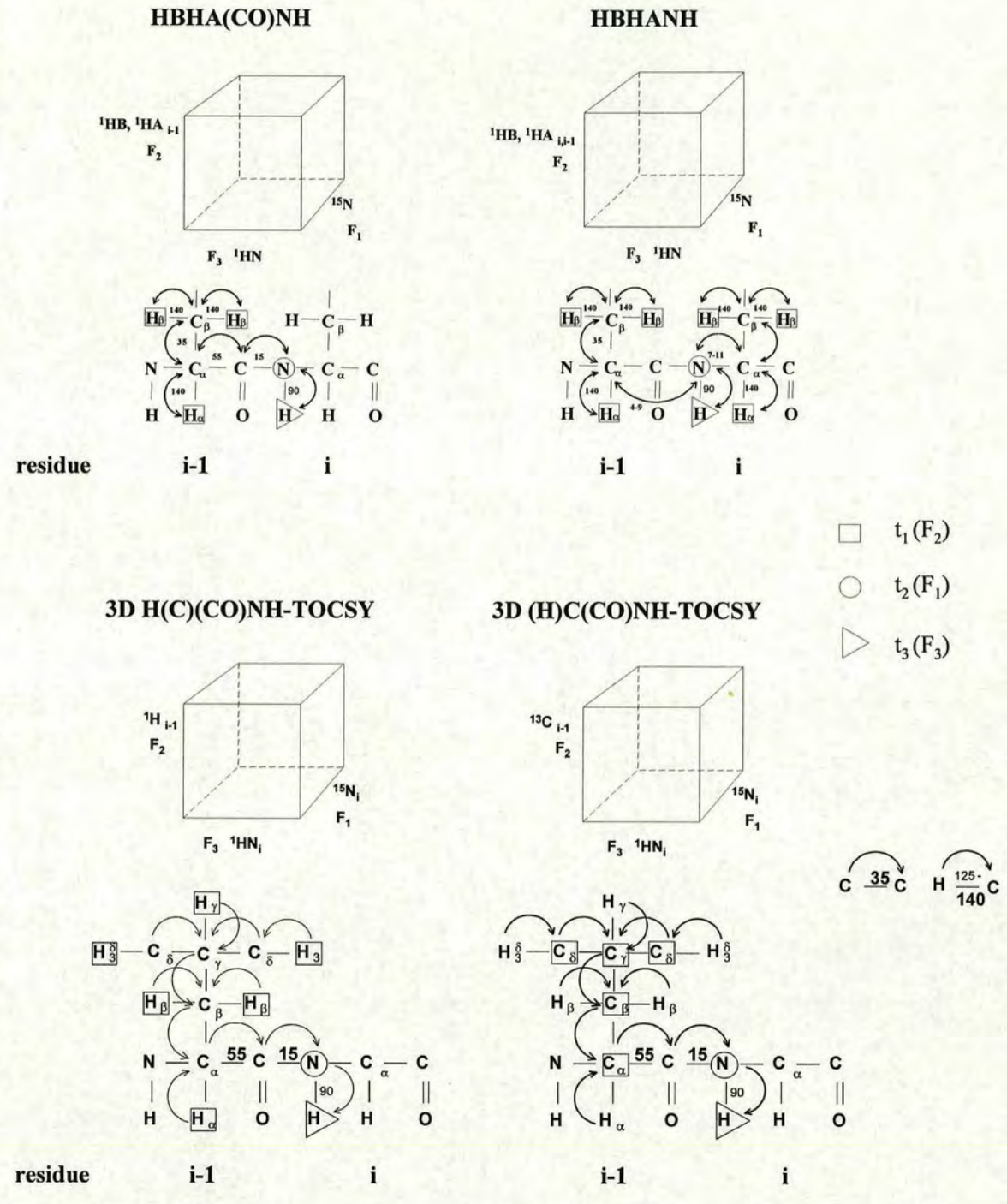



**Figure 2.23: Experiments for backbone and side-chain assignments in 15N/13C-labelled proteins**

Experiments are described as they are viewed in the Ansig software package for resonance assignments. Note: the so-called "out and back" implementation was used where the magnetisation starts and finishes on NH protons, as indicated by the double-headed arrows in the HBHA(CO)NH, and HBHANH experiment. In the TOCSY experiments, magnetisation does not start on NH protons.

# CHAPTER THREE: EXPRESSION AND PURIFICATION OF RECOMBINANT PROTEINS

## 3.1 Production systems for recombinant proteins

Both NMR and X-ray crystallography, require relatively large amounts of pure and homogenous material, 10-100 mg for a complete structural study [144]. Sample preparation is a "bottle-neck" of structural biology, and in the post-genomic aera this issue has become crucial. Structural genomics, i.e. the three-dimensional structure determination of all the protein folds in the entire proteome, will require efforts towards more efficient protein expression and purification systems, and possibly lead to their automation [145]. Currently obtaining sufficient amounts of pure protein can be a long, expensive and relatively complex process. This was emphasised in a recent study, which revealed that out of over one thousand proteins, derived from all three biological kingdoms, only 15-20% of small (< 50 kDa) non-membrane proteins were suitable immediately for structural biology [146]. The small number of proteins readily available for structure determination was compared to the "low hanging fruit" of structural genomics, and to some extent, this study showed also that many proteins cannot be expressed in soluble form in bacteria. Eukaryotic proteins often require extensive post-translational modifications or the presence of specific protein chaperones in order to fold and/or function correctly, and this renders bacteria unsuitable for their recombinant expression. The use of eukaryotic expression systems is therefore a good alternative to bacteria as they are likely to possess the correct cellular machinery to process such proteins and allow their soluble expression. Currently, these systems are predominantly yeast, insect cells and mammalian cells. Yeast is the most commonly used as it has some post-translational capacities, a short doubling time when compared to other eukaryotic systems, and is relatively inexpensive. One of yeast's main advantages is that metabolic labelling is still possible at reasonable cost, while labelling media for insect cells, although commercially available, is relatively expensive, as well as labelling media for mammalian cells. The issue of metabolic labelling is crucial to modern protein NMR. Besides the technical advantages it offers, isotope incorporation is highly desirable for high resolution NMR structures as the accuracy of protein structures determined by NMR is dependent on the extent of the data that one can obtain [147].

The time, the resources, the cost of protein production, and of course the level of expression achieved by a particular expression system are the parameters that one has to consider when undertaking any structural study. This often leads to a "trade-off" between expression systems. For instance, expression of one protein in an insect cells system might be achieved in soluble form but at a low level, while bacteria might achieve expression of the same protein in insoluble form, at a high level, which subsequently must be refolded.

This study involved the expression of extracellular proteins of human or rat sources, all of which contain established or putative disulphide bonds in various numbers, and which are often glycosylated. Thus, the methylotropic yeast *Pichia pastoris* was employed as an expression system. When levels of expression were low or undetectable in yeast, a bacterial system was used as an alternative.

### 3.1.1 Prokaryotic expression

*Escherichia coli (E. coli)* has long been the primary prokaryotic host for heterologous protein expression, therefore there is much available information and technology regarding this system involving many different recombinantly expressed proteins. The advantages of *E. coli* include its relatively easy, rapid and inexpensive methods of growth, transformation and maintenance. Furthermore, most of the biochemical pathways of *E. coli* are understood in great detail, and its entire genome has been sequenced. Foreign proteins can potentially be produced in *E. coli* in large amounts (5-50% of total protein). Nevertheless, prokaryotic cells such as *E. coli* are unable to perform many post-translational modifications, which occur in eukaryotic cells *(e.g.,* glycosylation, phosphorylation and disulphide bond formation), and which may be crucial for a foreign protein to adopt an active form. The expression and accumulation of a foreign protein in *E. coli* may also induce aggregates of protein [148, 149]. Sometimes the protein in these inclusion bodies can be refolded *in vitro* to produce functional protein, but this is not always possible and renaturation can be expensive and time consuming.

### 3.1.1.1 pET System

Various *E. coli* expression systems are commercially available, each system offering different benefits for protein expression and purification. All systems are designed to achieve a tightly controlled transcription coupled with a strong and rapid induction. Recently, new systems to enhance translation have been developed. The pET system is the best-characterised system that is commercially available (Novagen), and was employed in this study. pET-based vectors utilise bacteriophage T7 transcription signals to achieve very high levels of protein expression (Figure 3.1). The power of the pET system is that the T7 RNA polymerase is specific for its own promoter [150, 151], which is found only on the expression plasmid. Protein expression is achieved in an expression-*E. coli* strain that carries a chromosomal copy of the T7 RNA polymerase gene under the control of the lacI gene (DE3 strains). This theoretically enables T7 promoter repression during growth and culture expansion that otherwise might adversely affect bacterial growth. Some pET plasmids also carry the lacI gene and a lac operator sequence downstream of the T7 promoter (T7*lac*), to enhance repression of any basal expression. Other *E. coli* strains (pLysS and pLysE strains) have been engineered to contain a compatible plasmid that provides a small amount of T7 lysozyme during growth, which binds to T7 RNA polymerase and inhibits transcription. T7 RNA polymerase is expressed in the host cell only when induced with IPTG (isopropyl-beta-D-thiogalactopyranoside), a non-metabolizable analogue of lactose, allowing transcription from the T7 promoter. Induction is usually allowed to continue for a few hours.

### 3.1.1.2 Soluble versus insoluble and solubility tags

In inclusion bodies (see above) only a small fraction of the protein is folded correctly, and the rest is in a denatured and non-functional form. In some cases the formation of inclusion bodies allows purification of the recombinant protein by washing and pelleting [152].

Many vectors are now engineered to contain DNA sequences encoding a specific peptide or protein, which will be fused to the gene of interest and can theoretically be used for one-step purification of the recombinant protein by affinity binding. One of the first proteins to be exploited as a fusion

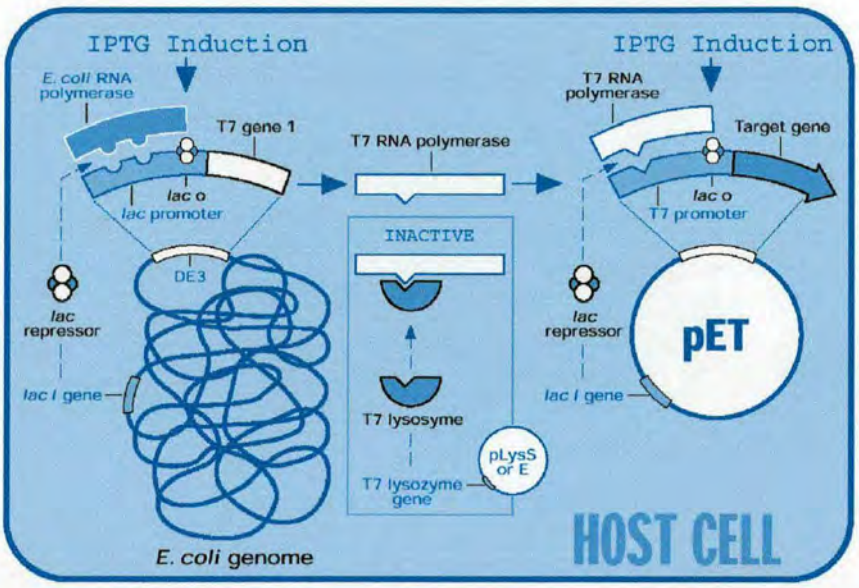



**Figure 3.1: pET system overview-*taken from http://www.novagen.com***

partner was glutathione S-transferase (GST) [153]. In vectors containing GST-fusion tag, the recombinant protein is fused with GST encoded by the parasite *Schistosoma japonicum* [151]. This allows purification of the fusion protein on a glutathione affinity Sepharose column. The "6xHis" Tag is also popular; in this system the recombinant protein is fused to a six-histidine peptide at its N or C-terminus. This latter is extensively used, especially due to its unique ability to work under denaturing conditions. In general, the recombinant fusion protein is engineered so that an endopeptidase cleavage site sequence is incorporated between the tag and the protein (e.g. thrombin, factor Xa, enterokinase, etc). Many tags have functions beyond purification; for example some of the traditional tags are stabilisers of protein folds (GST, Maltose binding protein) [151]. Recently tags with the sole purpose of enhancing solubility have been developed such as thioredoxin [154] or nusA tags [155].

Although *E. coli* lacks post-translational capacities, the use of ompT or ompA leader signal sequences can direct secretion of recombinant protein into the periplasmic space. This particular compartment has an oxidative environment and consequently is the only place where disulphide-bonded proteins have been observed in *E. coli*. Therefore targeting recombinant protein to the periplasmic space might lead to disulphide bond formation; such a strategy has been successful for the production of antibody fragments [151]. Recently, new tags for periplasmic localisation have been supplemented with the use of *E. coli* proteins that catalyse the formation (DsbA) and isomerisation of disulphide bonds (DsbC) [156].

### 3.1.2 Eukaryotic expression: the methylotrophic yeast *Pichia pastoris*

Mammalian cell expression systems are the best protein expression tools for mammalian proteins as they can produce native proteins, but they have extensive drawbacks in term of yield, mechanism of induction, and metabolic labelling. Use of non-mammalian hosts such as yeast or insect cells have several advantages. These expression systems have much better characterised genetics and biochemistry [157, 158], produce near-native glycosylation or at least some glycosylation, and have well defined secretory pathways for extracellular expression, and thus disulphide bond formation [159]. Yeast based expression systems are particularly attractive. Yeast can grow rapidly on minimal media, which enables metabolic labelling at reasonable cost. Expressed proteins can be engineered for cytoplasmic or secreted expression. Yeast has the ability to perform many eukayotic post-translational modifications. Finally, yeast is well suited for large-scale fermentation.

A highly sucessful system for the production of heterologous proteins is the methylotrophic yeast *Pichia pastoris* (*P .pastoris*) [160]. The basis of the *P. pastoris* expression system is methanol metabolism. Growth on methanol is mediated by alcohol oxidase, an enzyme whose *de novo* synthesis is tightly regulated by the alcohol oxidase promoter (AOX1). This enzyme has a very low specific activity; therefore to conpensate for this, the enzyme is overproduced, accounting for up to 30% of the total soluble protein of the cell [161]. There is a second alcohol oxidase which is encoded by a second AOX gene (AOX2), but has little specific activity and loss of the AOX1 gene lead to strains that exhibit poor growth on methanol medium, named $Mut^S$ strains (Methanol Utilisation Slow). Strains with both AOX genes correspond to the wild-type strain and are referred as $Mut^+$ strains (Methanol Utilisation plus) [161]. Invitrogen has developed plasmids that contain the AOX1 promoter to induce over-expression of recombinant proteins upon growth on methanol. Vectors containing the heterologous protein gene downstream of the AOX1 promoter are subsequently integrated into the *P. pastoris* genome by homologous recombination to maximise the stability of expression strains. The vector is designed to be linearised with a unique restriction site in either a marker gene (histidine gene) or the AOX1 promoter.

After transformation, the free DNA termini of the vector will recombine at the HIS4 locus or the AOX1 locus (depending on the restriction site employed-Figure 3.2), either by single crossing over (insertion) or double crossing over (replacement). Multiple insertions occur spontaneously shortly after transformation at a frequency of about 1-10% of the insertion events. Transformant

selection is achieved with the HIS4 selection marker, as engineered *P. pastoris* strains from Invitrogen are auxotrophic for histidine. Selection can also be performed using antibiotic resistance (zeocin or geneticin). Recombinant proteins are targeted for secretion by use of the α-factor leader sequence.

## 3.2 Materials and methods

This section describes briefly some of the essential biochemical techniques used in this study.

### 3.2.1 Standard molecular biology techniques

Polymerase chain reaction (PCR), and DNA manipulations were performed according to standard protocols [162]. The sequence integrity and the reading frame of all DNA constructs were confirmed by cycle sequencing.

### 3.2.2 Standard cell culture techniques

#### 3.2.2.1 *E. coli* work

*E. Coli* strains BL21(DE3) were purchased from Novagen in a chemically competent format and used for protein expression. For small-scale expression tests, a single colony from a fresh transformation was used to inoculate 5 ml of LB medium with antibiotic (100 μg ampicillin/ml). After growth and induction, 1.5 ml of culture was transferred to an Eppendorf tube, and spun at 13,000 g for 10 min at room temperature. The supernatant was decanted and the cell pellet was resuspended in 1 ml of Phosphate Buffer Saline (PBS) and spun again. For total cell protein extract, the cell pellet was resuspended in 1 ml of PBS, 100 μl was then mixed with an equal volume of 2 x SDS-PAGE loading buffer, and a sample of 15 μl was analysed by SDS-PAGE. For soluble fraction extract, the cell pellet was resuspended in 300 μl of Bug Buster reagent (Novagen) supplemented with 1 μl of Benzonase (25 units-Novagen), and incubated for 20 min on a rocking platform at room temperature.

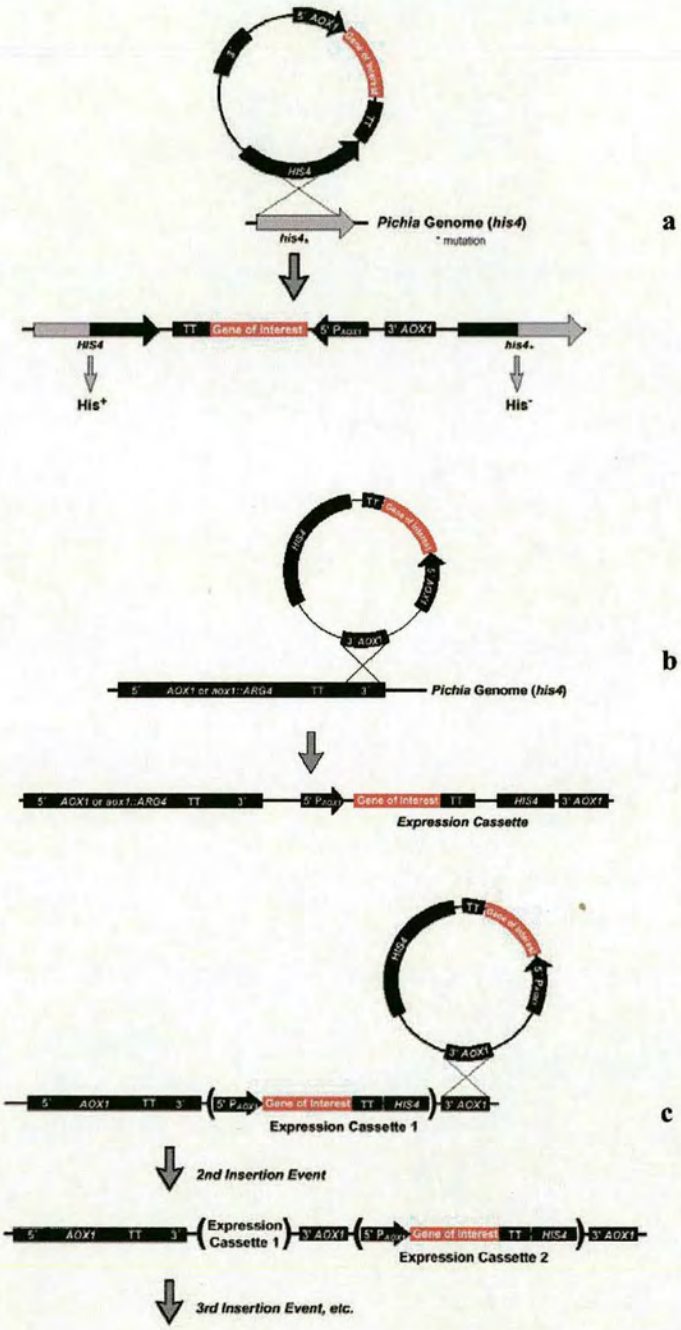



**Figure 3.2: Homologue recombination events of the *Pichia pastoris* expression vectors at different loci**

**a**) Gene insertion at the HIS4 locus. **b**) Gene insertion at the AOX1 locus. **c**) Example of multiple gene insertion. *Taken from reference* [161].

The cell debris was then pelleted by centrifugation (10 min, 13,000 g, room temperature), the soluble fraction decanted into a clean Eppendorf tube and an equal volume of 2 x SDS-PAGE loading buffer was added; a 30 µl sample was analysed by SDS-PAGE. For insoluble fraction extract, the cell pellet (above) was resuspended in 100 µl 2 x SDS-PAGE loading buffer, and a 15 µl sample was analysed by SDS-PAGE.

For large-scale bacterial cultures, a single colony from an overnight transformation was transferred to 100 ml of LB medium containing appropriate antibiotic and incubated at 37°C overnight in a shaking incubator (250 rpm). The next day, cells were harvested by centrifugation (2000 g, 5 min at room temperature), the supernatant was decanted and the cell pellet was resuspended in an equal volume of fresh medium containing antibiotic. The culture was then diluted 50-100 times (to a volume less than 25% of the flask volume), and the expression of recombinant protein was induced with 1 mM IPTG at $OD_{600}$ 0.7-0.9. Growth was continued for 3-4 hours at 37°C, and 250 rpm.

### 3.2.2.2 Yeast work

All yeast constructs were linearised and transformed in *Pichia pastoris* strain KM71 by the electroporation method [161]. Use of *Pichia pastoris* strain KM71 yields only $Mut^S$ transformants since this strain is itself exclusively $Mut^S$. Transformants were selected using zeocin and selected clones were tested for their capacity to secrete relevant fragments (see appendix A). Whenever possible, a clone apparently secreting the highest amount of recombinant protein was selected, and used for further large-scale growth to allow protein characterisation and authentication. *Pichia pastoris* strain KM71 has a mutation in the histidinol dehydrogenase gene (HIS4) [161], which prevents histidine synthesis. The pPICZ alpha plasmid [163] or its modified version used in this study does not complement HIS4 and selected transformants were further transformed with linearised pPIC9 vector to restore their ability to grow on histidine deficient medium. Alternatively, a pre-transformed pPIC9 *Pichia pastoris* strain KM71 was established and used directly in some transformations.

For sample preparation, cultures were grown in shake-flasks or high cell density fermentations. In each cases, the starter culture consist of a single colony selected from yeast peptone dextrose (YPD) Zeocin agar plates (see appendix A) and inoculated into 10 ml of buffered minimal glycerol medium (BMG) in a 50 ml Falcon tube, and incubated for 36 h at 30°C in a shaking

incubator (250 rpm). For shake-flasks cultures, this starter culture was used to inoculate 1 l of BMG, and split into four 1 l baffled flasks (Nalgene) capped with a BugStopper sterile closure (Whatman). Cultures were grown for another 36-48 h at 30°C (250 rpm) before harvest by centrifugation (1850 g for 5 minutes at room temperature) under sterile conditions. Cells were resuspended in one-fifth of the original culture volume (200 ml) of buffered minimal methanol medium (BMM) for induction. Induction periods were typically 5 days at 25-30°C (250 rpm) with daily addition of 100% methanol to maintain a final concentration of 0.5% (v/v).

For $^{15}$N-isotope labelling, cultures were also grown in shake-flasks, but $^{15}$N BMG and $^{15}$N BMM were used (see appendix A) which had the same composition as BMG and BMM media except that they contain 0.2% of $^{15}$N-ammonium sulfate (Goss Scientific-Cambridge Isotope labs) as the sole nitrogen source. For$^{15}$N/$^{13}$C-isotope labelling, three minimal media were used (see appendix A). $^{15}$N/$^{13}$C buffered minimal dextrose medium (BMD) that contain 0.2% $^{15}$N-ammonium sulfate, and 0.5% $^{13}$C-glucose (Goss Scientific-Cambridge Isotope labs) was used instead of BMG medium. Cells were then harvested by centrifugation as described above, and resuspended in $^{15}$N/$^{13}$C BMG that contains 0.2% $^{15}$N-ammonium sulfate and 0.1% $^{13}$C-glycerol for 2h prior to induction. This stage allows a smooth switch between glucose and methanol carbon source for the cells, and was found to prevent cell death. Induction was carried out using $^{15}$N/$^{13}$C BMM, which contains 0.2% $^{15}$N-ammonium sulfate and 0.5% $^{13}$C-methanol. To minimise protein degradation during production, the pH of both BMG and BMM solutions were lowered to 3, using phosphoric acid, as well as implementing a daily harvest and medium change. Quantitatively and qualitatively, no difference was observed between samples obtained from growth at pH 6 or 3.

In some cases, the shake-flask method in minimal media was replaced by a fermentation method for higher protein production. A starter culture as described above was used to inoculate 200 ml of BMG in a one-litre flask for 24 h ($OD_{600}$~6). This culture was then used to initiate a fermentor culture containing 2 l of *P. pastoris* basal salt medium and 6.7 ml of PMT1 trace metal salts (Invitrogen) contained within a 5 l working volume vessel on a BioFlo 3000 fermentation unit (New Brunswick Scientific Co). Dissolved oxygen concentration (DO) was maintained at 40%, the temperature at 30°C, the pH at 5 by using ammonium hydroxide, and antifoam (Sigma) was used to control excess foaming. After 18-24 h, DO increased to 100% indicating that glycerol was completely consumed and the glycerol fed-batch phase could be initiated. Cells were then fed with 50% w/v glycerol and 12 ml of PMT1 trace metal salts over 24 h at a rate of 40 ml per hour (18.15 ml/h/litre initial fermentation volume). At the end of the glycerol fed-batch step, again DO increased

to 100% indicating that glycerol was completely consumed. DO spikes were thus used to make sure the glycerol was limiting. Induction was then carried out for 3-4 days at a methanol feed rate of 7.5 ml per hour with an initial slow ramping to allow the culture to adapt smoothly to methanol.

### 3.2.3 Some general biochemical methods

General reagents were purchased from Sigma or otherwise stated. Buffers were prepared using Milli-Q grade water.

#### 3.2.3.1 Chromatography

Fast protein liquid chomatography (FPLC) was performed on a Waters 650E Advanced Protein Purification system fitted with a Waters 600 controller unit, a Waters 996 photodiode array detector and a Waters fraction collector. High-pressure liquid chromatography (HPLC) was carried out on a Waters system comprising a Waters 600 controller unit, a Waters 616 pump unit, a Waters 486 tunable absorbance detector and a Waters fraction collector. Alternatively, a BioCad 700E Perfusion chromatography workstation comprising a Scout Column selector and an Advantec SF-2120 super fraction collector was used for both FPLC and HPLC.

FPLC media for self-packed columns and pre-packed columns were from Amersham Pharmacia Biotech (ion exchange, hydrophobic interaction, gel filtration, and affinity chromatography) or from PerSeptive Biosystems (POROS media). HPLC pre-packed columns were from Brownlee and media for self-packed columns was from PerSeptive Biosystems (POROS media).

Self-packed Concanavalin A (Con A) sepharose and metal chelate affinity-chromatography columns were prepared with Amersham Pharmacia Biotech media and Novagen media, respectively. The media was packed in 20 ml columns from Biorad laboratories (Econo-Pac columns with 10 ml reservoir) and operated under gravity flow.

### 3.2.3.2 Denaturing electrophoresis of proteins

Protein samples were resolved by reducing SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using a Mini-PROTEAN II or Criterion gel system (Biorad Laboratories) and visualized using the Coomassie blue method. SDS-polyacrylamide gels were 15% acrylamide (self-poured). 8-16% and 10-20% SDS-polyacrylamide gradient gels were purchased in a precast format from Biorad Laboratories or Novex. Prestained molecular weight markers were from New England Biolabs.

In most cases, protein samples were precipitated by addition of trichlororacetic acid (TCA) solution to a 10% (v/v) final concentration prior to SDS-PAGE analysis.

### 3.2.3.3 Western blotting

Proteins were first resolved using SDS-PAGE before electrotransfer onto a nitrocellulose membrane (Amersham Pharmacia Biotech). Electro-transfer was performed in a Mini Trans-Blot cell (Biorad Laboratories) at constant voltage (150 mA) for 85 minutes in Towbin buffer (10% Methanol). Positive transfer was monitored using prestained molecular weight markers. Blotted membranes were blocked for 1 h with 5% Milk-phosphate buffered saline (PBS). Membranes were then washed over a period of 30 min with PBS (three buffer changes). Antibody incubation was performed at room temperature for an hour or overnight at +4°C. Detection of polyhistidine tags was carried out with anti-His6 monoclonal antibodies that are horseradish peroxidase conjugated (Sigma), and which allow detection without the need of a secondary antibody (1:800-1:500 dilution). The DAB (3,3'-diaminobenzidine) peroxidase substrate from Sigma was employed for staining.

### 3.2.3.4 Protein ultra-filfration and dialysis

Concentration of protein samples was achieved by ultra-filtration. Large volumes (from 1 l) were concentrated to 200 ml using a Millipore Prep/scale-TFF cartridge (regenerated cellulose) linked to a peristaltic pump (Millipore). Further concentration was performed with gas ($N_2$) -pressure stirred cells (Amicon, Millipore), which typically allowed concentration of several hundreds of ml to as little to 5 ml. Ultrafiltration membranes (Polyether sulfone) used in such stirred cells were purchased from

Pall Gelman Laboratory. Each recombinant protein sample was concentrated using an appropriate molecular weight cut-off. Final concentration, in the range of several hundreds of microliters, was achieved by centrifugal ultra-filtration with centricon concentrators (Millipore) or Vivaspin concentrators (Sartorius). Dialysis tubing for buffer exchange or use in refolding protocols was purchased from Pierce and Warriner (Spectra/por6).

### 3.2.3.5 Estimation of protein concentration

Protein concentration was calculated using absorbance at 280 nm and a theoretical extinction coefficient based on the protein sequence (ProtParam Tool, Expasy Tools –http://www.expasy.ch).

### 3.2.3.6 Mass spectrometry analysis

Ion electrospray mass spectroscopy (atmospheric pressure) was carried out on a Micromass Platform II single quadrupole mass spectrometer. Samples were prepared in 50% acetonitrile-water, 0.5% formic acid. Positive ionisation was normally used.

### 3.2.3.7 N-terminal sequencing

For protein N-terminal sequencing, protein samples were first resolved using SDS-PAGE and electro-blotted on PVDF (Polyvinylidine difluoride) membranes (ProBlott-Applied Biosystems). Membranes were stained with 0.1% Coomassie blue (G-250) in methanol and destained with 10% acetic acid-50% methanol. Relevant bands were submitted for protein N-terminal sequencing on a ABI 477A sequencer at the Welmet protein characterisation facility (University of Edinburgh).

## 3.3 Results and discussion

### 3.3.1 Preliminary remarks

Successful protein expression for structural biology often requires the testing of different systems and hosts. Protein purification can also be problematic and may require modifications of the original construct derived from the cDNA to fine tune domain-boundaries or simply to aid the purification process by using affinity tags. Such investigations are to some extent based on trial-and-error, and may be extremely time-consuming. This work illustrates the difficulties in defining domain-boundaries when undertaking the structural study of modular proteins.

Yeast has been the host of choice throughout this study, since all constructs are, or at least suspected to be disulphide bond containing as well as N-glycosylated. CP1 and FTM2 expressions were of particular interest. CP1 NMR data exhibited poor $^{1}H$ chemical shift dispersion and lack of crosspeaks in the $^{15}N$-HSQC spectrum, indicating that it may be poorly structured, aggregated or in dynamic exchange (see chapter four). As mentioned earlier, YTH identified the C-terminus of the fibulin protein family as a ligand for the complement module pair in the $GABA_B$ receptor 1a. Early YTH screens localised the interaction region to the last three epidermal growth factor-like (EGF) modules (repeat 8, 9, and 10) and the FTM2. Further screens refined this interaction to the FTM2 on one hand, and CP1 on the other, which consequently led to a focus on the FTM2 only. Nevertheless the array of EGF modules (EGF~8-10) were expressed and purified, and subsequently used to raise antibodies against fibulin-2.

### 3.3.2 Yeast expression

#### 3.3.2.1 $GABA_B$ R1a complement module pair constructs

The poor quality of the NMR data of the recombinantly expressed $GABA_B$ R1a complement module pair, led to an investigation of the influence of the factors that could be important for proper folding. Complement module sequences are characterised by the presence of four cysteine residues that form the two disulphide-bonds of a module and in general also define the module boundaries. The $GABA_B$ R1a N-terminal sequence does contain the consensus sequence for two complement modules, but there are seven residues N-terminal to the first cysteine of the first module that include a

potential N-glycosylation (Asn23; Figure 3.3-a). This first predicted module also includes a second consensus motif for N-glycosylation (Asn83), as well as an extra tryptophan residue in the predicted first loop. An overview of the different constructs that have been made and examined by NMR spectroscopy is presented in Figure 3.3-b. These constructs are presented, in the following sections, in the chronological order in which they were made and follow the investigation of the influence of the N-terminal extension in CP1 on protein folding. The role of N-glycosylation was investigated also through mutagenesis and by means of endoglycosidases. All yeast expression constructs used the *Pichia pastoris* pPICZα A expression vector (Figure 3.3-c).

#### 3.3.2.1.1 TR-CP12

The $GABA_B$ R1a "truncated" complement module pair (TR-CP12) corresponds to residues 24-159 in the original $GABA_B$ R1a rat protein sequence (numbering includes the protein signal sequence). This protein fragment is of the same sequence as in humans except for two conservative mutations that are found in the second predicted complement module: glutamic acid 128, and valine 144 are replaced by aspartic acid and isoleucine, respectively.

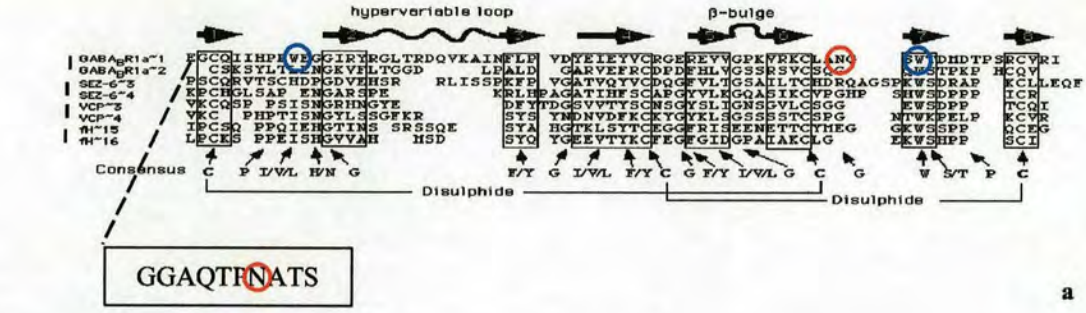
A yeast strain, expressing the TR-CP12 was prepared by Dr. Edward Hawrot in our laboratory [34]. The yeast clone was shown to produce two forms of the recombinant fragment, as seen by SDS-PAGE (Figure 3.4-a). The major protein in the sample has a mass of ~16 kDa while the minor form has a mass of ~20 kDa. Mass spectroscopy confirmed a mass of 15628.7 for the major form, which corresponds, to the expected mass of the recombinant protein fragment if all eight cysteines are in disulphide bonds, as would be expected for two complement modules, and with an EAEF N-terminal sequence from cloning, but without an N-glycan. The minor form was found to have the same N-terminal sequence and to correspond to an N-gylcosylated form of the major protein. This minor protein can be treated by endoglycosidase (EndoH) and led to a band shift by SDS-PAGE, consistent with molecular weight of the major form [34].

Following Dr. Hawrot's work in our laboratory, the current work started with large-scale protein preparations for structure determination. Typically, preparations were on a 1 l BMG-200ml BMM scale, as described previously, and yielded up to ~10 mg of protein. The 200 ml of culture obtained at the end of the induction period (5 days) was harvested, and spun at high speed (10000 g, 30 min, 4°C), the supernatant was isolated and protease-inhibitors were added (1 mM PMSF/10-20 mM EDTA). Preparations were kept at 4°C prior purification, alternatively stored at –20°C. Protein

concentration was carried out by ultrafiltration using 3 kDa molecular weight cut-off membranes, to a volume of 20 ml. Originally, the protein sample was purified according to Dr Hawrot's protocol using a 1 ml reverse phase C18 column equilibrated in 90% $H_2O$-10% acetonitrile with 0.1% v/v trifluoroacetic acid (TFA). Purification was achieved by an elution gradient from 80% $H_2O$ 20% acetonitrile [0.1% v/v TFA] to 60% $H_2O$ 40% acetonitrile [0.1% v/v TFA] at a flow rate of 1 ml/min over 34 column volumes. This allowed only a partial resolution of the two protein forms; the minor (N-gylcosylated) form was shown to elute slightly earlier than the major form. This result was in agreement with Dr. Hawrot's findings when he repeated the same experiment in his own laboratory at Brown University (RI-USA).

Subsequently, a new purification strategy was devised. Several ion exchange chromatographic conditions were investigated, and both cation and anion exchange were tested. Best conditions were found to be as follow: the concentrated protein was first desalted against 50 mM sodium acetate-pH 5.3 using gel filtration (PD-10 column), and purified by cation exchange chromatography on a 1-ml MonoS column, equilibrated in the same buffer. Elution was achieved with a gradient of 0-100% 1 M NaCl in buffer (above) at a flow rate of 1 ml/min over 25 column volumes. This allowed a initial capture of the two forms, which eluted as a single peak (Figure 3.4-b), as shown by SDS-PAGE. Fractions containing protein were pooled and concentrated to 20 ml. The two protein forms were then resolved by hydrophobic interaction chromatography as follows: the protein concentrated from cation exchange chromatography was first desalted against water and then an equal volume of 0.2 M sodium phosphate buffer, pH 7, containing 3.6 M ammonium sulfate was added. The protein was applied to a 1-ml phenyl superose column equilibrated in 0.1 M sodium phosphate buffer, pH 7 containing 1.8 M ammonium sulfate. Elution was achieved with a gradient of 100-0% ammonium sulfate at a flow rate of 0.5 ml/min over 25 column volumes (Figure 3.4-c).

The peak containing the minor form eluted earlier than that of the major form (Figure 3.4-c&d). Fractions from between the peaks containing a mixture of the two forms were discarded and each form was concentrated separately using Vivaspin6 concentrator (5 kDa cut-off). Buffer-exchange against an appropriate NMR buffer, and concentration of protein, were achieved simultaneously by several rounds of concentrations and dilutions (typically, 5-6 rounds of 10 times concentration-dilution) to a final volume of 300-600 µl.



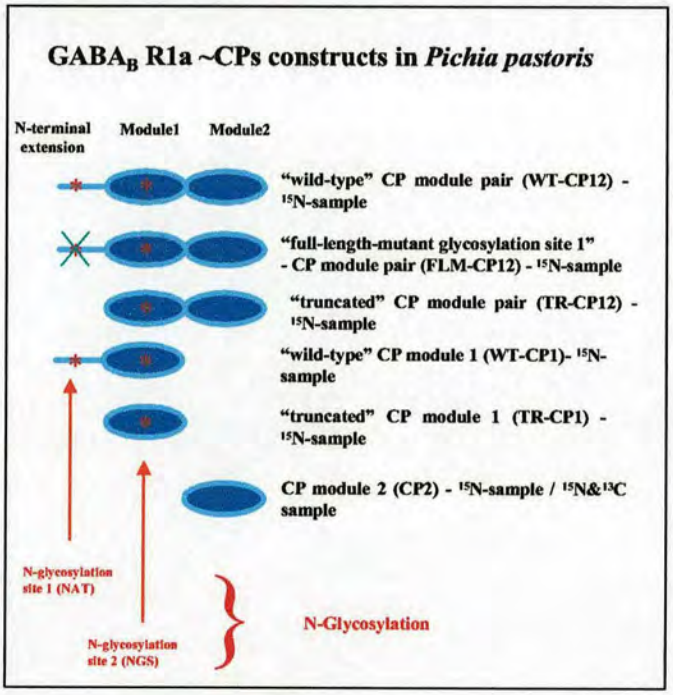



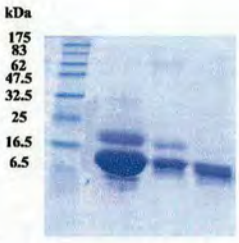


**Figure 3.3: Overview of the GABA$_B$ R1a complement modules expression in *Pichia pastoris***

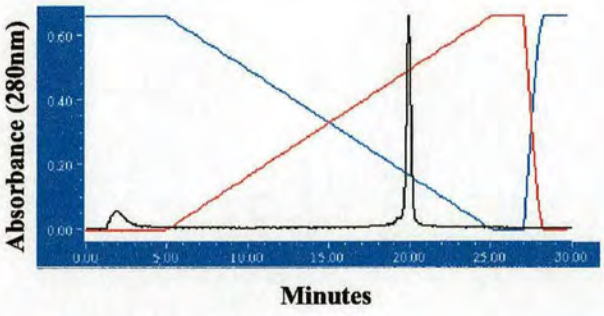
**a)** Protein sequence of the predicted complement control protein (CCP or CP) modules against others characterised modules [34]. An alignment of the GABA$_B$ R1a pair of CP-modules with a CP-pair found in the CNS (the 3rd and the 4th CPs from the rat SEZ-6 protein), a CP-pair found in the VCP viral protein (2$^{nd}$ and 3$^{rd}$ modules), and a well characterised CP-pair found in the complement system (the 15th and the 16th CPs from the human factor H protein).

**b)** Summary of the protein constructs used in this study.

**c)** The pPICZα vector from Invitrogen and its multiple cloning site.

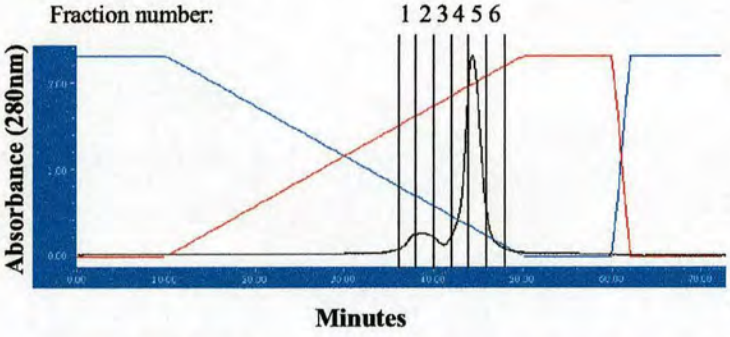



a



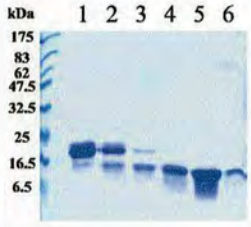

**50 mM sodium acetate / pH 5.3**
**50 mM sodium acetate / pH 5.3 + 1 M NaCl**

b




**0.1 M sodium phosphate / pH 7 + 1.8 M ammonium sulfate**
**0.1 M sodium phosphate / pH 7**

c




d

**Figure 3.4: Purification of TR-CP12**

**a)** Crude preparation concentrate separated by SDS-15% PAGE under reducing conditions, and visualized using the Coomassie blue method. One ml of a 200 ml preparation was TCA precipitated.

**b)** Cation exchange chromatogram (blue: buffer, red: eluent).

**c)** Hydrophobic interaction chromatogram (blue: buffer, red: eluent).

**d)** Hydrophobic interaction chromatogram fractions separated by SDS-15% PAGE under reducing conditions, and visualized using the Coomassie blue method. Lanes 1-3: 250 µl of fractions 1-3, TCA precipitated. Lanes 4-6: 50 µl of fractions 4-6, TCA precipitated.

For mass spectroscopy, samples were desalted using a reverse phase C4 column (eluting with an acetonitrile gradient identical to that used for the C18 column, mentioned above), freeze-dried, dissolved in 50% acetonitrile in aqueous solution containing 0.1% TFA and analysed by electrospray-ionisation mass spectroscopy.

Two compounds with masses of 15628.5 Da and 15428.3 Da were found to be contained within the major form. These correspond to: the expected molecular weight of TR-CP12 plus the cloning artefacts (EAEF) at the N-terminus for the 15628.5 Da compound (expected mass is 15628.7 Da); and the same but with cloning artefacts (EF) only at the N-terminus for the 15428.3 Da compound (expected mass is 15428.5 Da). These two products represent alternatively processed products of the α-factor leader sequence. Incomplete removal of the EA spacers in α-factor pre-proteins is very common in yeast [158]. The measured masses are also consistent with expected molecular weights if all eight cysteines are in the disulphide form. The presence of cloning artefacts (EAEF) at the N-terminus was confirmed by N-terminal sequencing. A third compound of 14892.6 was observed in some samples (Figure 3.5) and corresponds to the expected molecular weight of the TR-CP12 with the first five N-terminal residues missing (expected is 14892.95 Da). Missing N-terminal residues in recombinant proteins expressed in yeast has been reported and seem also to be the result of incorrect cleavage of the pre-protein. A form missing only the first two N-terminal residues (15152.2 Da) has also been observed by Dr. Hawrot [34].

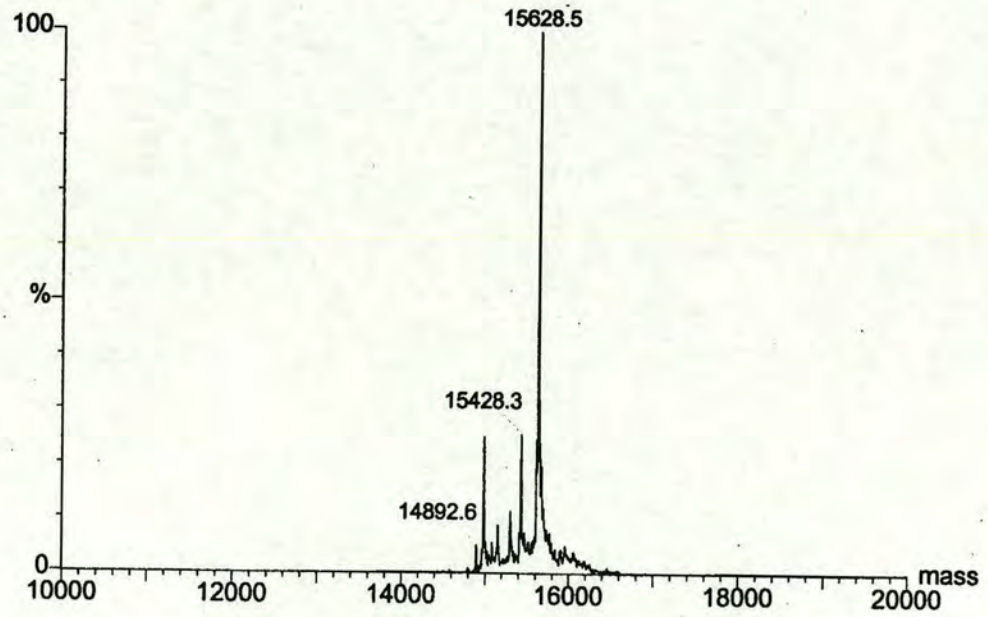



**Figure 3.5: Mass spectrum of the major form of TR-CP12**

The minor form mass spectrometric analysis was of poor quality, and not reproducible. Protein purity was greater than 95% by SDS-PAGE, but despite extensive desalting, a mixture of compounds appeared in the mass spectrum of molecular weights between 17.3 and 18.6 kDa: 17360 Da, 17492 Da, 17816 Da, 17978 Da, 17994Da, 18322 Da. Such molecular weights are consistent with the ~20 kDa band seen by SDS-PAGE, and seem to correspond to heterogeneous N-glycosylation at asparagine 83, with glycan chains that differ in their number of mannose units. The 17492 Da compound represents the 15628 Da compound found for the major form with nine mannose units and two N-acetylglucosamine (GlcNAc) units, similarly the 17816 Da compound represents an adduct of 11 mannose units and two GlcNAc, and the 17978 Da compound represents an adduct of 12 mannose units and two GlcNAc. Such heterogeneity could also explain the lack of reproducibility of mass spectrometry measurements between different preparations.

#### 3.3.2.1.2 FLM-CP12

The $GABA_B$ R1a "full-length" complement module pair, with the N-glycosylation site 1 (Asn~23) mutated out (FLM-CP12) corresponds to residues 17-159 in the original $GABA_B$ R1a rat protein sequence. This region encompasses the two predicted complement control modules with the native seven-residue N-terminal extension described above. This extension contains a putative N-glycosylation consensus motif (NAT), which could result in covalent glycan linkage at asparagine 23. To facilitate downstream purification, this site was engineered (N23A) by PCR to abolish the consensus motif.

The DNA encoding for TR-CP12 was obtained from Dr. Hawrot, and was used as a PCR template to amplify FLM-CP12. The 5' primer (see appendix B) contained the sequence encoding the seven-residue extension with asparagine 23 changed to an alanine, as well as an *EcoRI* site. The 3' primer contained a *NotI* site. The amplified fragment and the vector pPICZα A (Invitrogen) were each double digested with the restriction enzymes *EcoRI* and *NotI* (37°C-overnight), agarose-gel purified and finally ligated by incubation with T4 ligase (16°C-overnight).

The ligation product was transformed into calcium chloride competent-Top 10F *E. coli* cells (Invitrogen), and positive transformants were screened by restriction digest after plasmid DNA minipreparation. The 5' and 3'AOX1 sequencing primers were both used in separated sequencing reactions, to confirm both the authenticity and the frame of the cloned insert in the pPICZα A vector.

A sample of the DNA construct (50 μg) from a midi-preparation was digested with *SacI*, to generate a linear construct. Linearized DNA was phenol-chloroform extracted once, ethanol precipitated and resuspended in sterile distilled water. A sample containing 10 μg of linear DNA construct in 10 μl was used for *Pichia pastoris* transformation by the electroporation method (KM71 strain). Positive transformants appeared after two days on YPDS zeocin plates (see appendix A), 20 transformants were screened for protein expression (small-scale protein production test – see appendix A for method). One transformant expressing good levels of recombinant fragment, as judged by SDS-PAGE analysis, was selected and transformed with the linearized (*SacI*) pPIC9 vector (Invitrogen) as described previously, to replace the HIS4 gene lacking in the KM71 strain. Transformants were selected on minimal dextrose (MD) plates (see appendix A), and one or two positives were further screened using a small-scale test as described above. Large-scale preparations and protein purification were as described for the TR-CP12, and yielded similarly up to 10 mg of recombinant protein. The FLM-CP12 was also expressed as two forms, a minor N-glycosylated form and a major non N-glycosylated form (Figure 3.6).

Mass spectrometry analysis of the major form revealed a single compound of 16010 Da, which is consistent with the expected molecular weight of the FLM-CP12 plus the cloning artefacts (EF) at the N-terminus (expected mass is 16011.12 Da). Such a mass is also consistent with the expected molecular mass as if all eight cysteines are in the disulphide form. The presence of the cloning artefacts (EF) at the N-terminus was confirmed by N-terminal sequencing.

#### 3.3.2.1.3 WT-CP12

The $GABA_B$ R1a "wild-type" complement module pair (WT-CP12) corresponds to residues 17-159 in the original $GABA_B$ R1a rat protein sequence. The seven-residue extension as described above is included but now has the original protein sequence containing the N-glycosylation consensus motif.

The DNA construct encoding FLM-CP12 obtained above was used to amplify the DNA encoding for WT-CP12 by PCR. The 5' primer (see Appendix B) contained the sequence coding for the native seven-residue extension with asparagine 23, as well as an *EcoRI* site. The 3' primer contained a *NotI* site. Cloning procedure and yeast transformation were as described for the FLM-CP12. Large-scale preparations were also similar to the FLM-CP12.

The recombinant WT-CP12 was expressed as two forms, but the yield was lower than that obtained for previous constructs. Following deglycosylation by the endoglycosidase EndoH$_f$, the two forms collapsed to a single band by SDS-PAGE, consistent with the expected molecular weight of the non N-glycosylated protein form, with intensity now similar to the previous constructs. This evidence for hyper- and heterogeneous N-glycosylation at asparagine 23 and is consistent with observations of other N-glycosylated proteins overexpressed in this study (see EGF~8-10 expression, section 3.3.2.3.2). Hyper N-glycosylation in yeast has been described in many published studies [158], and has been observed by many other researchers in our laboratory (Andrew Herbert - unpublished, Dr. Joanne O'Leary - PhD Thesis). Subsequently, all WT-CP12 samples were exhaustively deglycosylated by use of endoglycosidase Endo H$_f$.

Protein purification was achieved by cation exchange chromatography approximately as described for the previous fragments, although cation exchange condition varied slightly: binding buffer and elution buffer were 12.5 mM sodium acetate, pH 5.3 instead of 50 mM sodium acetate, pH 5.3. After cation exchange, samples were concentrated, and again buffer-exchanged against the cation exchange binding buffer prior to deglycosylation with 2500 units of endoglycosidase EndoH$_f$ (4 h, 37°C) per mg of recombinant protein. Endoglycosidase Endo H$_f$ was subsequently removed by reloading the cleavage mixture on cation exchange column, and further purification was achieved by reverse phase chromatography (C8 column, same gradient as described previously). SDS-PAGE analysis and NMR spectroscopy confirmed removal of the N-glycans, although in some cases the recombinant protein was further purified on a Concavalin A (Con A) sepharose column to remove uncleaved N-glycosylated forms prior to reload on the cation exchange column. For Con A sepharose chromatography, the protein sample was desalted by gel filtration (Hiprep_Desalting column) against water, an equal volume of 2x stock Con A binding buffer (40 mM Tris-HCl, pH 7.4, 1 M NaCl) was added, and the protein sample was applied to a 5 ml Con A sepharose column (gravity flow) equilibrated in the same buffer (1x: 20 mM Tris-HCl, pH 7.4, 0.5 M NaCl). The protein sample was loaded, 2 ml at a time, to allow binding of uncleaved N-gylcosylated forms, the column was then washed with five column volumes of binding buffer, and absorbance at 280 nm was used to monitor the protein content of each 2-ml fraction. Protein yields were in the order of 8 mg per preparation.

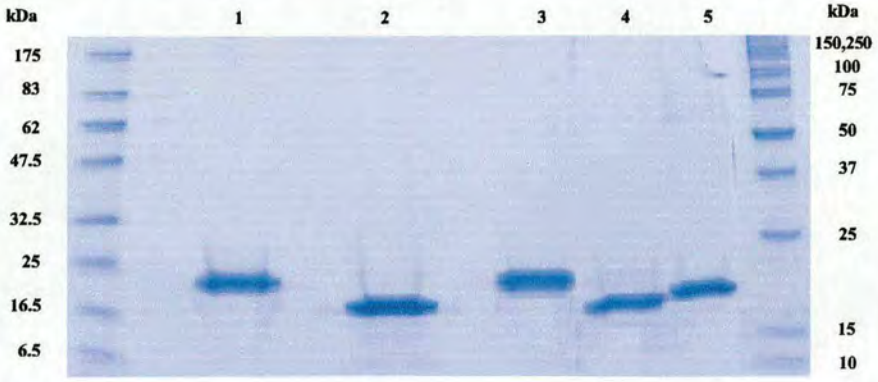
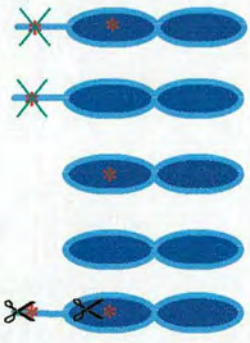



**Figure 3.6: SDS-(10-20%) PAGE analysis of the $GABA_B$ R1a complement module pair fragments under reducing conditions**

Each lane corresponds to ~20 μg; visualised using the EZ blue stain (Sigma- Coomassie blue method).



Lane 1: FLM-CP12 glycosylated

Lane 2: FLM-CP12 non glycosylated

Lane 3: TR-CP12 glycosylated

Lane 4: TR-CP12 non glycosylated

Lane 5: WT-CP12 deglycosylated.

Mass spectrometry analysis revealed four compounds with molecular weight of 16056.6 Da, 16072.6 Da, 16259.3 Da and 16275.6 Da. The first mass corresponds to the expected mass of the WT-CP12 plus the cloning artefacts (EF) at the N-terminus (expected mass is 16054.14 Da) plus 2.5 Da. The second compound corresponds to the first one plus an unknown adduct of 16 Da. The third compound corresponds to the first one with an adduct of 202.7 Da, which probably corresponds to one GlcNAc unit (average mass is 203.2 Da). The last compound corresponds to the second one with an adduct of 203 Da, corresponding to one GlcNAc unit. The presence of the cloning artefacts (EF) at the N-terminus was confirmed by N-terminal sequencing.

The molecular weight found for the first compound is therefore 2.5 Da out when compare to the expected weight for the WT-CP12. Although such a difference could account for breakage of one disulphide bond, this seems unlikely as in the other fragments the mass analysis indicates all cysteines are disulphide bonded. This discrepancy could be the result of protein modification(s) that occur on some unidentified residues; one possibilty is deamination of asparagine or glutamine to aspartate or glutamate, which in both case account for the again of one Da. The adduct of 16 Da could be explained by hydroxylation (of lysine or tryptophan), or more likely by oxidation of methionine. A discrepancy of +14-16 Da was also found in the module 1 constructs, where it will be shown that such modification is not related to disuphide bond formation (see below).

### 3.3.2.2 $GABA_B$ R1a CP single module constructs

#### 3.3.2.2.1 CP2

The $GABA_B$ R1a complement module2 (CP2) corresponds to residues 96-159 in the original $GABA_B$ R1a rat protein sequence. The putative "linker" sequence (residue 96-residue 98) was included at its N-terminus. The DNA encoding for CP2 was amplified from the truncated module pair DNA construct (see Appendix B) and cloned with a similar directional cloning strategy. Yeast transformation and large-scale preparations were as described previously.

CP2 was also purified by cation exchange chromatography. The concentrated protein was first desalted against 50 mM sodium acetate, pH 4.6 using gel filtration (PD-10 column), and purified by cation exchange chromatography on a 1-ml MonoS column equilibrated in the same buffer. Elution was achieved with a gradient of 0-100% 1 M NaCl in buffer at a low rate of 1 ml/min over 25 column volumes. Protein samples were buffer-exchanged, and concentrated as described previously but membranes with a 1 kDa molecular weight cut-off were used for the concentrators. Sample was of great purity according to SDS-PAGE analysis (Figure 3.7-a) and mass spectrometry (Figure 3.8).

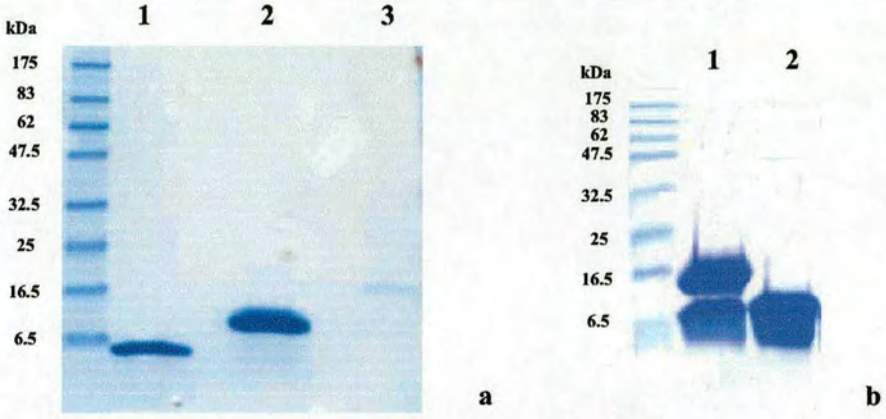
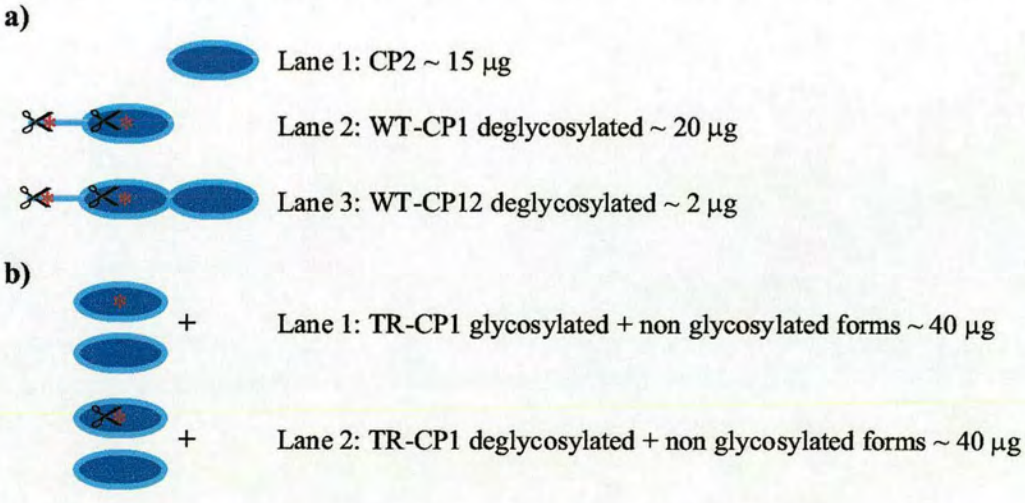



**Figure 3.7: SDS-(10-20%) PAGE analysis of $GABA_B$ R1a complement module fragments under reducing conditions, visualised using the EZ blue stain (Sigma- Coomassie blue method)**



Mass spectrometry analysis of some samples revealed a single compound of 7482 Da, a molecular weight consistent with the expected molecular weight of CP2 fragment plus the cloning artefacts (EAEF) at the N-terminus (expected mass is 7481.42 Da) with all cysteines disulphide bonded. In some samples, two additional compounds of 7152 Da and 7005 Da were observed (Figure 3.8), which correspond to species missing N-terminal residues. The 7152 Da compound is missing the three N-terminal residues (expected mass is 7152.11 Da), while the 7005 Da compound is missing four N-terminal residues (expected mass is 7004.93). The presence of the cloning artefacts (EAEF) at the N-terminus was confirmed by N-terminal sequencing.

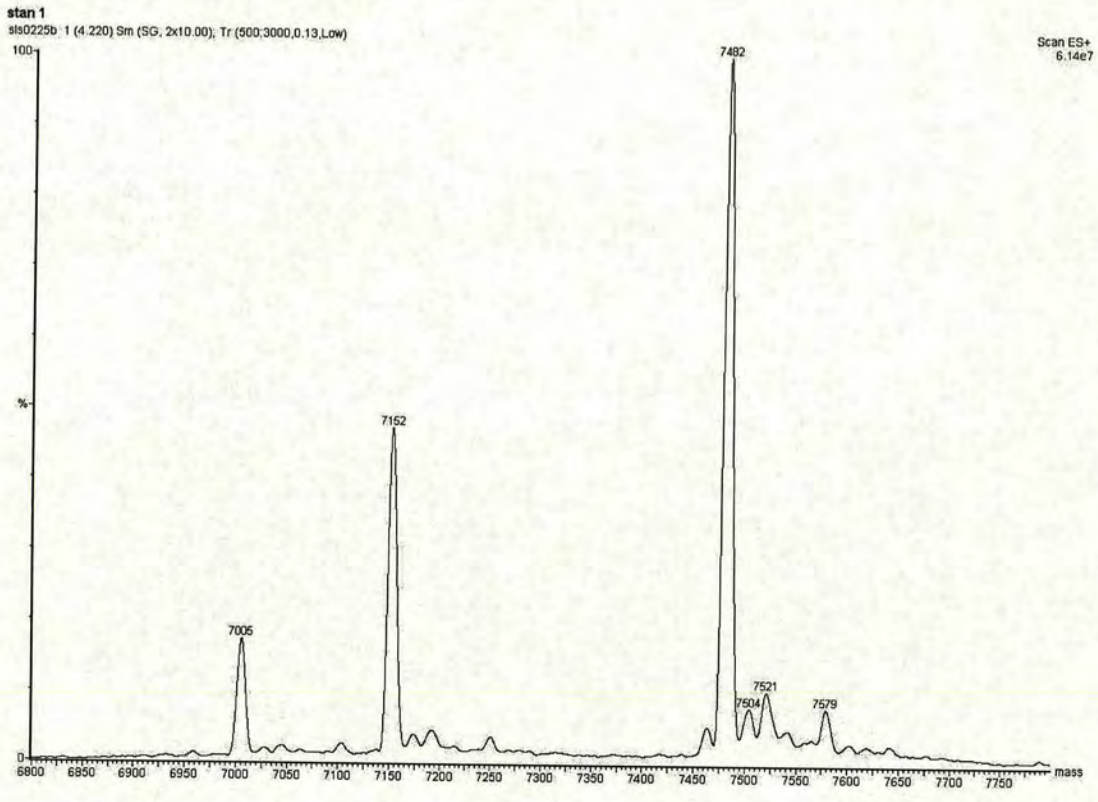



**Figure 3.8: Mass spectrum of CP2**

#### 3.3.2.2.2 TR-CP1

The $GABA_B$ R1a "truncated" complement module1 (TR-CP1) corresponds to residues 24-98 in the original $GABA_B$ R1a rat protein sequence. As for the TR-CP12, the N-terminal seven-residue extension is absent but "linker" sequence (residue 96-residue 98) between the two putative CPs is included.

The DNA encoding for the TR-CP1 was amplified from the TR-CP12 DNA construct (see Appendix B) and cloned with a similar directional cloning strategy. Yeast transformation and large-scale preparations were as described previously. Protein purification was similar to the WT-CP12 (including deglycosylation), except cation exchange buffers were 25 mM sodium acetate, pH 5.3. Membranes with a 3 kDa molecular weight cut-off were used for the concentrators. Sample gave a single band by SDS-PAGE analysis (Figure 3.7-b).

Three species with masses of 8809 Da, 9011 Da, and 9026 Da were found for the major form. These correspond to the expected molecular weight of the truncated module1 plus the cloning artefacts (EF) at the N-terminus for the 8809 Da compound (expected mass is 8810.06 Da – cysteines oxidised), and the cloning artefacts (EAEF) at the N-terminus for the 9011 Da compound (expected mass is 9010.26 Da – cysteines oxidised). These two compounds represent alternatively processed products of the α-factor leader sequence. The third compound corresponds to the 9011 Da compound with an adduct of 15 Da. The presence of the cloning artefacts (EAEF) at the N-terminus was confirmed by N-terminal sequencing.

#### 3.3.2.2.3 WT-CP1

The $GABA_B$ R1a "wild-type" Complement module1 (WT-CP1) corresponds to residues 17-98 in the original $GABA_B$ R1a rat protein sequence. As in the WT-CP12, the seven-residue extension with the N-glycosylation consensus motif was included. This construct also contains the entire putative "linker" sequence (residue 96 – residue 98) between the two predicted modules.

The DNA encoding WT-CP1 was amplified from the WT-CP12 DNA construct (see Appendix B) and cloned with a similar directional cloning strategy. Yeast transformation was as described previously but the *Pichia pastoris* KM71 strain was transformed with pPIC9 linearised vector before use to avoid the necessity of retransformation for histidine complementation. Methods for large-scale preparation and protein purification were similar to those used for the $GABA_B$ R1a

WT-CP12. Protein samples were buffer-exchanged against appropriate NMR buffer by several rounds of concentrations and dilutions as described previously with 3 kDa molecular weight cut-off membranes. The sample gave a single band by SDS-PAGE analysis (Figure 3.7-a).

Analysis by mass spectrometry revealed three major compounds with molecular weights of 9453 Da, 9657 Da, and 9860 Da (Figure 3.9). The first species corresponds to the expected molecular weight of the WT-CP1 (cysteines oxidised) plus the cloning artefacts (EF) at the N-terminus (expected mass is 9435.7 Da), and an adduct of 17.3 Da. The second compound correspond to the first one plus an adduct of 204 Da, which corresponds to one GlcNAc unit (average mass is 203.2 Da). The last compound correspond to the second one with an adduct of 203 Da, corresponding to a second GlcNAc unit.The discrepancy of 17.3 Da in the first species is close to 18 Da, and could correspond to hydrolysis of a peptide bond with the resulting fragments held together by a disulphide bond, for example. This hypothesis can be ruled out from experiments shown in section 3.3.2.2.4, and it is more likely to correspond to an adduct of 16 Da as seen in the module pair constructs, since mass spectrometer calibration can be out by 1 Da. The presence of the cloning artefacts (EF) at the N-terminus was confirmed by N-terminal sequencing.

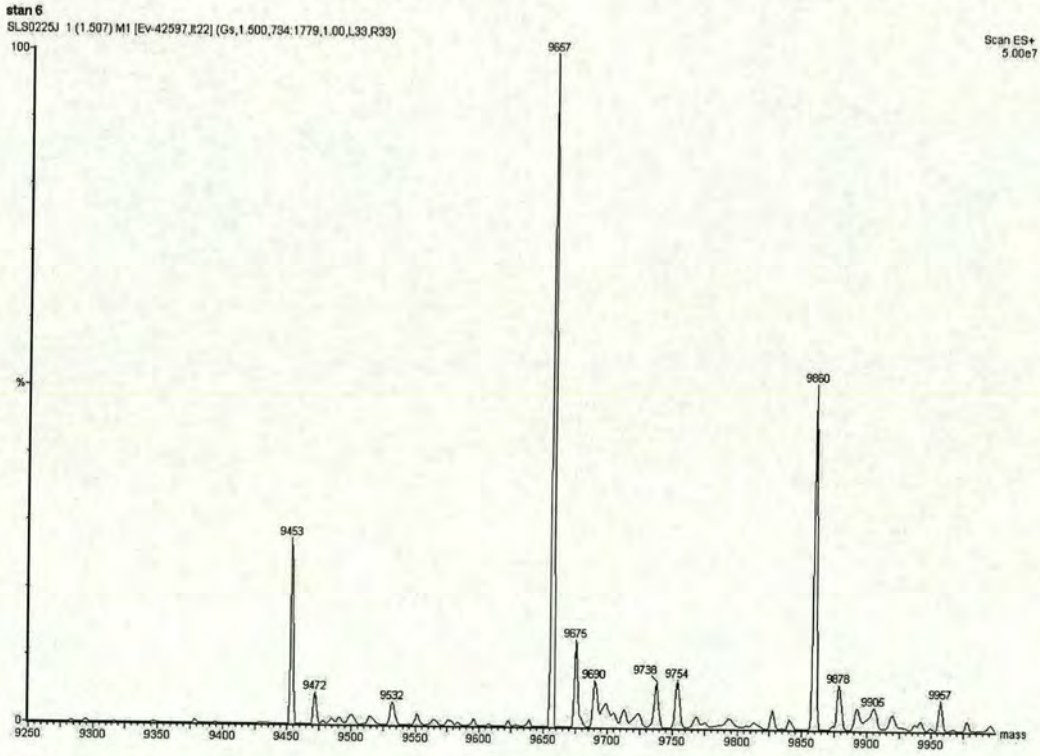



**Figure 3.9: Mass spectrum of WT-CP1**

#### 3.3.2.2.4 Disulphide bond pattern in the $GABA_B$ R1a CP1 constructs

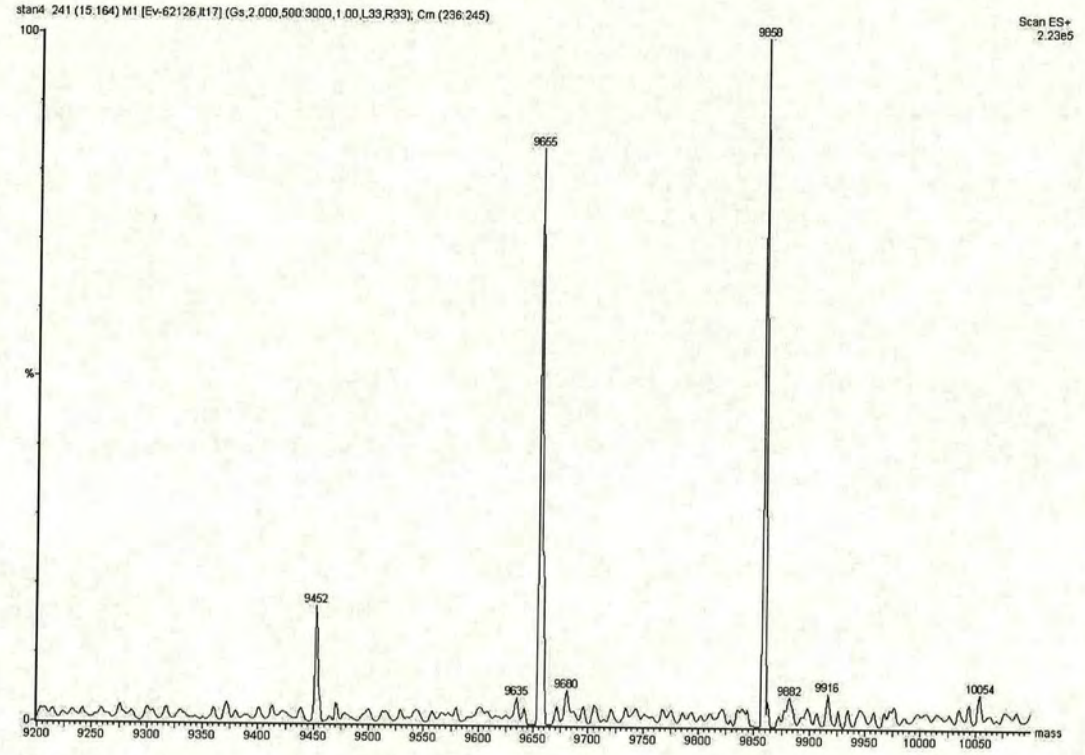
##### 3.3.2.2.4.1 Deduction of the number of disulphide bonds by reduction and carboxymethylation

The discrepancies in the mass analysis of the CP1 constructs raise the question of proper disulphide bond formation, and correct chain length. Reduction of disulphide bonds using DTT and subsequent carboxymethylation of free cysteines by neutralised iodoacetic acid allows deduction of disulphide bonds in protein, and identification disulphide bonded proteolytic fragments.

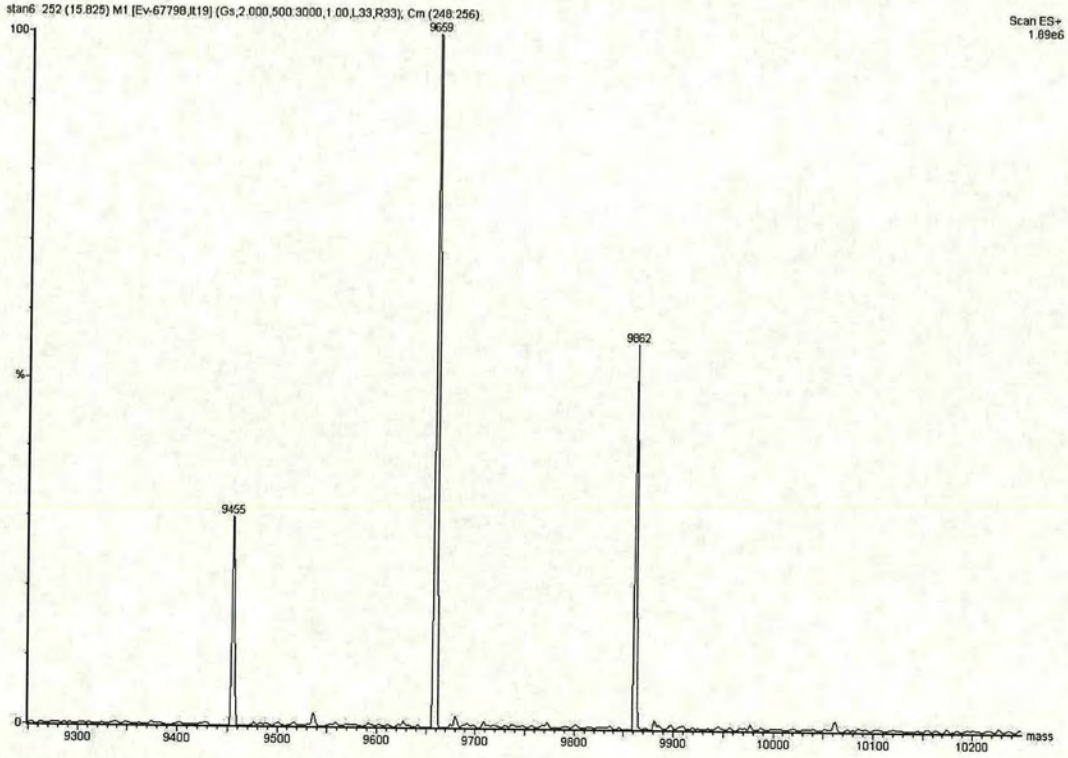
A sample of WT-CP1 (100 μg in 100 μl of 0.5 M Tris-HCl, pH 8.6, 5 mM EDTA, and 8 M urea), was reduced with 10 mM DTT for 2 h at 40°C under $N_2$. A second sample of WT-CP1 that has been denatured and reduced as above was further carboxymethylated with neutralised-iodoacetic acid (1.44 M stock in 1 M NaOH solution-final concentration was 40 mM) for 15-30 min at room temperature in the dark. In parallel, a third sample of WT-CP1 (100 μg in 100 μl of sterile distilled water) was carboxymethylated with neutralised iodoacetic acid under the same conditions as above but without prior denaturation or reduction. Finally, a fourth sample, in water, was reduced with DTT as above.

All four samples: denatured-reduced, denatured-reduced-carboxymethylated, carboxymethylated only, and reduced only were analysed by liquid chromatography mass spectrometry (LCMS). The resulting mass spectra are shown in Figure 3.10 and Figure 3.11.

Figure 3.10-a shows the mass spectrum of the reduced only form - some mass changes can be observed when compared to the original mass spectrum shown in section 3.3.2.2.3. The three species previously observed were 9453 Da, 9657 Da, and 9860 Da. After reduction with DTT, masses for the same sample are 9452 Da, 9655 Da, and 9858 Da. The apparent decrease of 1-2 Da observed for each species must correspond to the reduction of at least one disulphide bond, although the expected mass difference for complete reduction is 4 Da. Once "denatured and reduced" (Figure 3.10-b), all three species showed an adduct of 3-4 Da when compared to the reduced-only species (9455 Da, 9659 Da, and 9862 Da), which is consistent with the reduction of one or both disulphide bonds. It seems therefore that denaturation is required for reduction of disulphide bonds.




**a**




**b**

**Figure 3.10: LCMS of WT-CP1**

**a)** Reduced only

**b)** Denatured, and reduced

Interestingly a discrepancy of 4-5 Da is observed between the species from the carboxymethylated-only control sample (9451 Da, 9654 Da, and 9858 Da; Figure 3.11-A) and the respective species from the "denatured and reduced" sample. This is consistent with the complete reduction of both disulphide bonds under the latter conditions. No smaller fragments were observed following disulphide reduction, which implies that the peptide chain of the WT-CP1 is intact.

Figure 3.11-a show the mass spectrum of the carboxymethylated-only form; no significant mass changes can be observed when compared to the original mass spectrum shown in section 3.3.2.2.3. Following carboxymethylation, but without disulphide reduction, peaks at 9451 Da, 9654 Da, and 9858 Da are obtained. The apparent decrease of 2-3 Da observed for each species is of unknown origin but may well arise from calibration error between the two experiments; but interestingly, this time, lead to an adduct of 15.3 Da instead of 17.3 Da when comparing the first species of the carboxymethylated-only to the expected mass (9435.7 Da – cysteines oxidised). Carboxymethylation by iodoacetic acid is expected to occur specifically on free thiol groups, although side-reactions involving tyrosine, tryptophan and histidine have been reported . This type of modification introduces a $CH_2COOH$ group on free thiol, which would result in an adduct of 58 Da. No adduct of 58 Da can be detected on any of the three species and therefore it can be deduced that no free thiol groups are exposed in the native WT-CP1. Once denatured-reduced, and carboxymethylated (Figure 3.11-b), each of the three species originally observed in the carboxymethylated-only sample showed an adduct of 237-235 Da. Such adducts correspond closely to four times the expected 58 Da adduct. It can therefore be deduced from this last experiment and from the previous one that both disulphide bonds are formed in native WT-CP1, and the polypeptide chain is intact. This implies that the adduct of 15-17 Da observed in the WT-CP1 is not due to peptide cleavage, which, anyway, would add 18 Da. The adduct is most likely a result of methionine oxidation (only one methionine in the CP1 amino-acid sequence), or lysine hydroxylation.

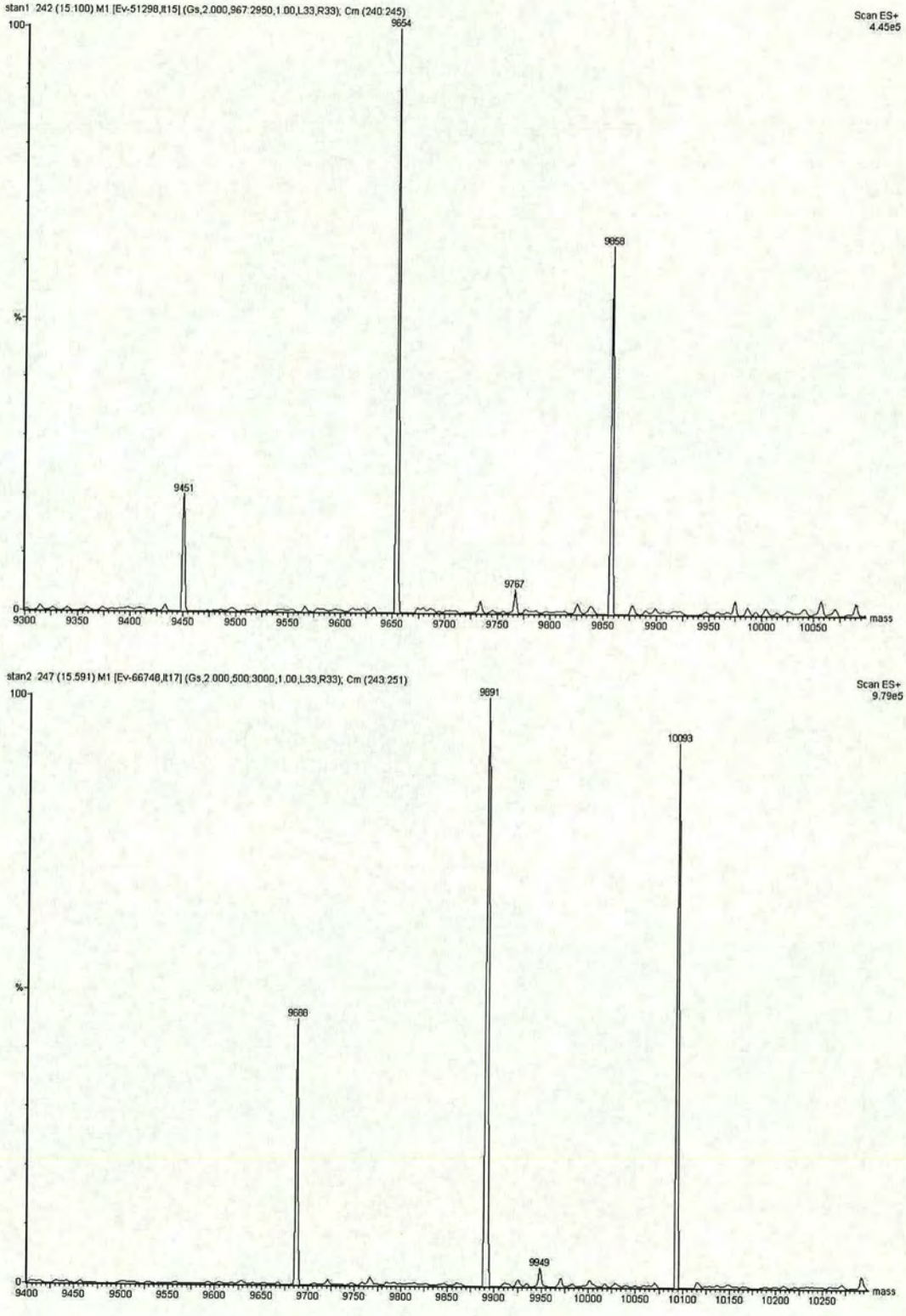



**Figure 3.11: LCMS of WT-CP1**

**a)** Carboxymethylated only

**b)** Denatured, reduced, and carboxymethylated

#### 3.3.2.2.4.2 Endoproteases digestions and their fragments analysis by Q-TOF Mass spectroscopy

Two types of endoproteases were used: Trypsin and *S. aureus* V8 Protease. Digestions and mass spectrometry analysis on Q-TOF (Quadrupole Time Of Flight) mass spectrometer were performed by Dr. Adam Gouldsworthy (Edinburgh University). Digests were performed on the TR-CP1 construct. Only trypsin digestion was successful, and this clearly identified the presence of the disulphide bond involving cysteine II (Cys130) and IV (Cys156). Most of the fragments were from non-specific cleavage reactions and were identified by their amino acids sequences. Some of these fragments are shown in Table 3.1.

**Table 3.1**

| Fragment mass (Da) from Trypsin digest (cut after R&K) | Corresponding amino acid sequence & observed sequence (underlined) |
|---|---|
| **1159.63 ($[M+2H]^{2+}$)** | **(K)AINFLPVDYEIEYVCR(G) + (R)CVR(I)** |
| **674.01 ($[M+3H]^{3+}$)** | **(N)FLPVDYEIEYVCR(G) + (R)CVR(I)** |
| **643.31 ($[M+2H]^{2+}$)** | **(Y)EIEYVCR(G) + (R)CVR(I)** |

Taken together, denaturation/reduction - carboxymethylation experiments and endoprotease digestions strongly suggest the presence two disulphide bonds for the $GABA_B$ R1a CP1. Cysteines II and IV are paired as expected, which obviously implies a cysteine I and III pairing. Consequently, it can be concluded that the disulphide bond found in the CP1 constructs are as expected for a CP module. Carboxymethylation experiments were also performed on WT-CP12, and confirmed the absence of free cysteines, the presence of four disulphide bonds and an adduct of 16 Da (data not shown).

### 3.3.2.3 Fibulin-2 constructs

The recombinant expression of fibulin-2 fragments identified from yeast-two-hybrid (YTH) screens (see chapter one) was undertaken to provide material for both structural and binding studies. DNA coding for different the fibulin-2 constructs were designed on the basis YTH results and their subsequent cloning in a *Pichia pastoris* expression vector was performed by Dr. Julia White (GlaxoSmithKline – Stevenage - UK). All DNA constructs (Figure 3.12) were cloned in a modified version of the commercially available vector pPICZα A vector from Invitrogen. It allows fusion of the targeted protein fragment with a c-myc epitope tag and a hexahistine tag, but was engineered in our laboratory (Dr. Nick Mullin and Andrew Herbert - unpublished) to include a thrombin-specific cleavage site to allow their subsequent removal.

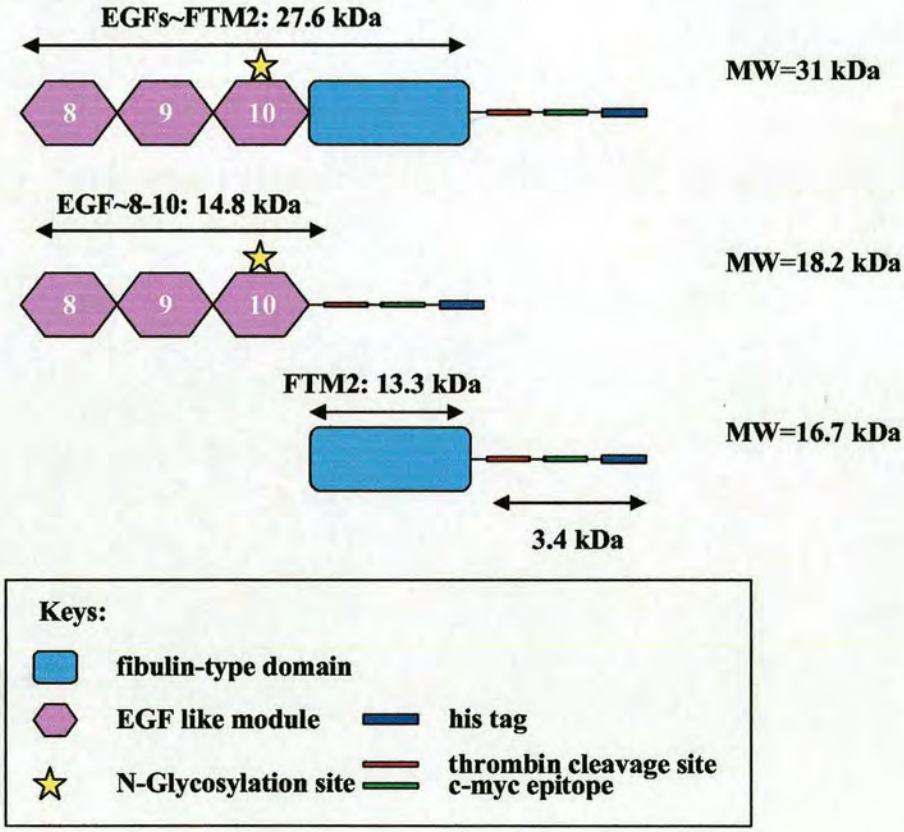



**Figure 3.12: Fibulin-2 constructs in *Pichia pastoris***

#### 3.3.2.3.1 Fibulin-2 EGF like repeats 8-10 and fibulin type module

The fibulin-2 epidermal growth factor like repeats 8-10 and fibulin-type module (EGFs~FTM2) construct correspond to the C-terminal end of fibulin-2 (Swiss Prot accession number P98095), and more specifically to residues 942-1184 in the original fibulin-2 human protein sequence (numbering includes the protein signal sequence). This fragment of fibulin-2 contains three consensus sequences for epidermal growth factor like modules (residues 942-1072) and the recently identified fibulin-type module (residues 1069-1184) shared with fibrillins. One N-glycosylation site is found at position Asn1035 in the third putative EGF like module.

The DNA encoding for EGFs~FTM2 was cloned in the modified pPICZα A vector. The resulting plasmid was then linearised (*SacI*) and yeast was transformed as described for the $GABA_B$ R1a fragments. Transformation yielded to a large number of transformants, and ~30 were screened in induction tests with Endo $H_f$ treatment prior to SDS-PAGE analysis. All failed to show any expression of the recombinant fragment. Subsequently, some transformants were grown and induced on a larger scale (100 ml BMG, 20 ml BMM); then their supernatant was concentrated, and treated with Endo $H_f$. Again no expression could by detected by SDS-PAGE (Figure 3.13), or by Western blot (anti-His6 antibodies, not shown).

The reason for the lack or the very low expression of EGFs~FTM2 is unknown. A large number of transformants were present, and other protein constructs (that later were found to express protein) were transformed using the same *P. pastoris* preparation, thus providing a control for the transformation. Cell pellets were also assayed by Western blotting (anti-His6 antibodies-not shown), but also failed to show any expression. Interestingly a faint band at ~25 kDa is present in the SDS-PAGE analysis from the larger scale preparation mentioned above, and could represent either a host protein or alternatively EGFs~FTM2 free from the *c-myc* epitope and His6 tag; which therefore could account for the lack of detection by Western blotting. Time didn't allow further investigation of such a possibility, nor further characterisation of the EGFs~FTM2 yeast clone(s). Western blotting with anti *c-myc* antibodies as well as PCR analysis to check construct integration into *P. pastoris* genome could be the focus of future work. However, expression in *P. pastoris* is not always successful, and there is 50-75% chance of expressing a protein in *P. pastoris* [161]. One explaination for no expression in *P. pastoris* is premature transcriptional termination - AT base repeats present in the DNA construct might ressemble the consensus sequence for such a transcriptional termination [161]. However, no such a sequence can be readily identify in the DNA coding for EGFs~FTM2.

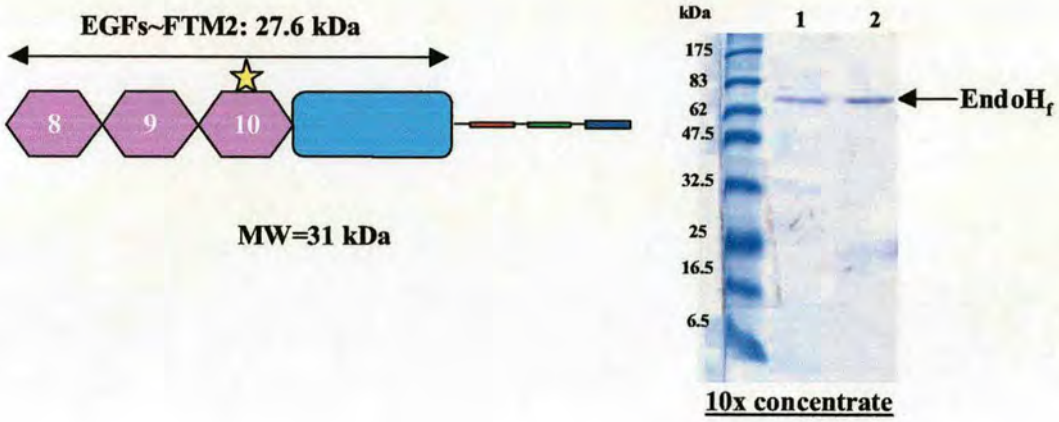



**Figure 3.13: Expression tests of the EGFs~FTM2 construct**

Two clones were grown (100 ml BMG) for two days, then harvested and induced for 5 days (20 ml BMM, 0.5% methanol daily). Supernatants were concentrated to a final volume of 2 ml using 5 kDa molecular cut-off concentrators. For each clone, one ml of supernatant was treated with Endo Hf, analysed by SDS-15% PAGE under reducing conditions, and visualized using the Coomassie blue method. Lane 1 is clone #10, and lane 2 is clone #12.

Note: cartoons legend as in Figure 3.12

#### 3.3.2.3.2 Fibulin-2 epidermal growth factor like repeats 8-10

The fibulin-2 epidermal growth factor like repeats 8-10 (EGF~8-10) construct (residues 942-1072) was transformed as described in the case of the $GABA_B$ R1a fragments. All transformants were shown to express the fragment at a good yield (Figure 3.14-a). The fragment was heavily glycosylated, and couldn't been detected without prior Endo $H_f$ treatment. This could be the result of heterogeneous hyper-glycosylation combined with the low susceptibility of glycosylated protein to TCA precipitation. Figure 3.14-a also illustrates clonal variation, i.e. how the level of recombinant protein production differs from one clone to another. Such variation arises from multiple insertion of the DNA construct encoding for EGF~8-10. Large-scale preparation of the EGF~8-10 construct was as described in the case of the $GABA_B$ R1a fragments. After induction, the culture supernatant was concentrated to 20 ml by ultra-filtration using 3-5 kDa molecular weight cut-off membranes, and Endo $H_f$ treated (6h, 37°C).

The protein concentrate was then desalted against 20 mM sodium phosphate, pH 7.4, 0.5 M NaCl using gel filtration (PD-10 column), and purified by immobilized metal affinity chromatography (Figure 3.14-b) on a HiTrap metal chelating column (1 ml) charged with nickel and equilibrated in the same buffer. Elution was achieved with a 0-100% 0.5 M imidazole gradient in buffer over 25 column volumes. Sample was of good purity according to SDS-PAGE analysis, and ~ 8 mg were buffer-exchanged against 1xPBS buffer, concentrated with concentrators of a 5 kDa molecular weight cut-off membrane, and submitted for antibody-production to our collaborator, Dr. Jeffrey McIlhinney (Oxford University-MRC Neuro-anatomy Unit). However protein characterisation was proven difficult, as the EGF~8-10 construct failed to be seen by mass spectrometry despite repeated attempts. Finally N-terminal sequencing reveal the expected N-terminal sequence (first eight residues with the cloning artefacts EF).

#### 3.3.2.3.3 Fibulin-type module from fibulin-2

The fibulin-type module from fibulin-2 (FTM2), residues 1069-1184, was transformed as described for the EGFs~FTM2 construct. Induction tests failed to show any recombinant expression, and larger scale growths also failed. Nevertheless, tags used for protein detection are putatively removable and it was hypothesised that a combination of proteolytic degradation and low expression level might account for the lack of detection by Western blot. Consequently, a fermentor was used to provide larger quantities of material.

A positive yeast transformant for the FTM2 was grown in a benchtop fermentor (BIOFLO 3000- New Brunswick scientific). The fermentation followed a standard protocol (2 l), including a glycerol batch feed (1 l of 50% glycerol) and a methanol batch feed for induction (100 h). Fermentation samples were taken every 24 h. None of these sample contained material for a direct visualisation of FTM2 overexpression. After completion, the fermentation supernatant (3 l) was concentrated to 20 ml. Overexpression was still undetectable. In an attempt to isolate any small amount of protein that could have been produced, the concentrated supernatant was fractionated using a different type of chromatography (not shown). The FTM2 construct includes a fused hexa-histidine tag, therefore nickel affinity chromatography was used to try to isolate some protein, but only host proteins were found in the bound fraction. Cation exchange chromatography allowed isolation of a protein that could correspond to the fibulin type module construct free of tags but N-terminal sequencing revealed the sequence of a host protein.

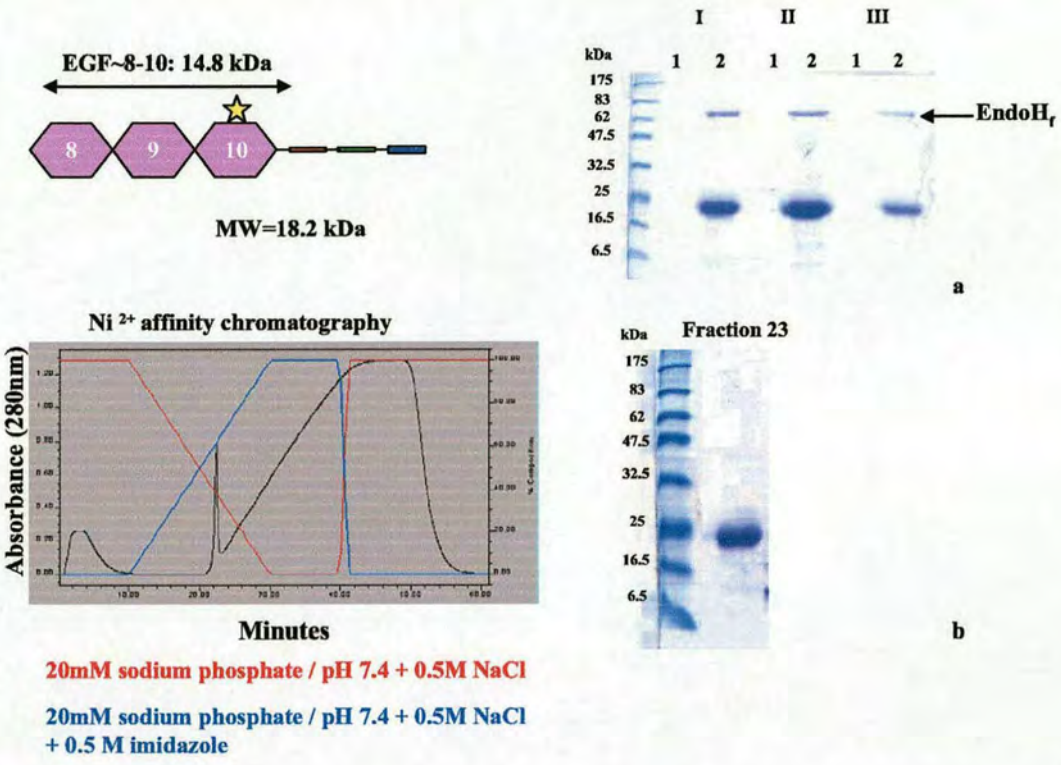



**Figure 3.14: Expression of the EGFs~FTM2 construct and EGF~8-10**

**a)** Expression tests of the EGF~8-10 construct: three clones (I-III) were analysed by SDS-15% PAGE under reducing conditions, and visualized using the Coomassie blue method. For each clone, lane1 corresponds to the crude supernatant and lane 2 corresponds to the same crude supernatant pre-treated with Endo $H_f$. Each lane is one ml of the expression test supernatant TCA-precipitated.

**b)** Nickel-affinity chromatogram of the large-scale preparation of the EFG~8-10 construct (red: buffer, blue: eluent). Also shown is the SDS-15% PAGE analysis under reducing conditions of the major peak from the nickel-affinity chromatography (fraction 23 min, one ml fraction were collected every minute). 100 µl of fraction 23 was TCA-precipitated. Gel was SDS-15%, and visualised using the Coomassie blue method.

Note: cartoons legend as in Figure 3.12

### 3.3.3 Bacterial expression

The lack of expression of the FTM2 in yeast led to the investigation of alternative hosts. The FTM2 (residues 1069-1184) was therefore cloned into prokaryotic expression vectors to allow recombinant expression in *E. coli*. As mentioned above, successful recombinant protein expression in *E. coli* is frequently achieved in the case of cytosolic proteins. However, extracellular proteins often accumulate as inclusion bodies. The FTM2 was expressed in various pET systems; only when fused to a thioredoxin solubility tag was the protein fragment recovered in a soluble form.

#### 3.3.3.1 Fibulin-type module from fibulin-2 as inclusion bodies

Cloning of the FTM2 in pET15b (Novagen; Figure 3.15) results in fusion with a hexa-N-terminal histidine tag (Figure 3.16). This allows the use of immobilized-metal affinity chromatography for downstream purification of the recombinant protein. This tag is cleavable by use of a specific protease, thrombin. The DNA construct encoding for the FTM2 in the pPICZα A-derived vector mentioned above was used as a PCR template. The 5' primer (see appendix B) contained an *NdeI* site. The 3' primer contained an *XhoI* site. Cloning of the PCR product in pET15b was performed as described previously, in a non-inducible *E. coli* strain (TOP10). For protein production, BL21(DE3) chemically competent cells were transformed with the FTM2-pET15b according to Novagen's protocol. Small-scale expression tests and large-scale growth for protein production were as described previously.

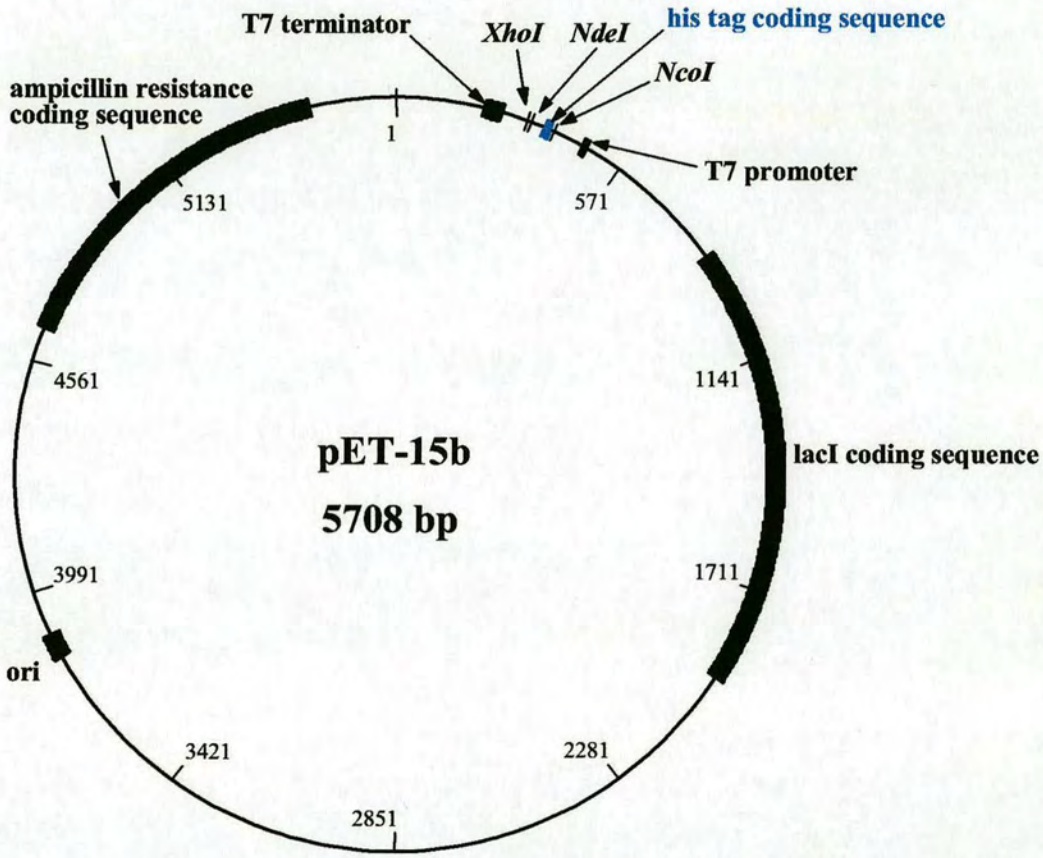



**Figure 3.15: pET15b vector from Novagen**

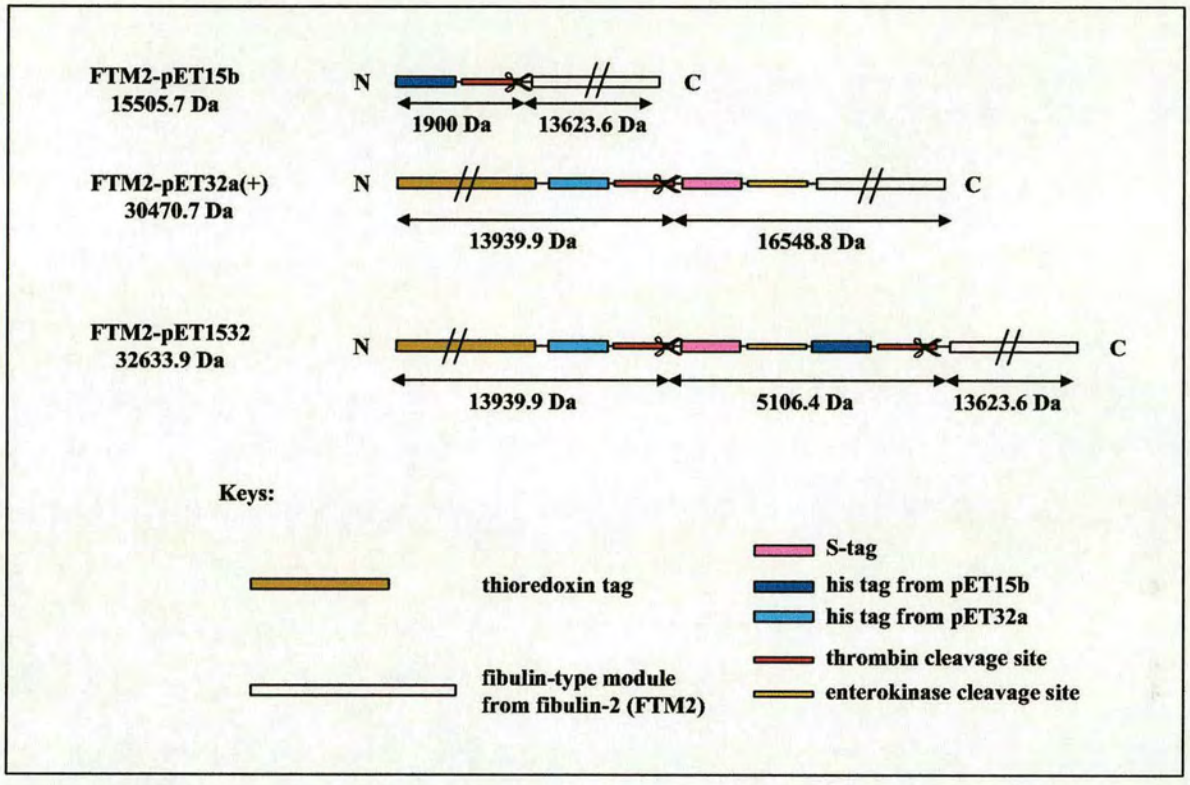



**Figure 3.16: Overview of the *E. coli* expression**

After 4 h of induction (1 mM IPTG), cells were harvested by centrifugation (6500 g, 4°C, 15 min), washed in PBS, and centrifuged again. The cell pellet was typically 3.6 g per litre of growth medium. Cell extract was prepared by use of the Bug Buster reagent (28 ml, Bug Buster is a mild detergent from Novagen) and benzonase (700 units), incubated for 20 min on a rocking platform at room temperature with protease-inhibitors ("Complete Protease inhibitor cocktail tablets without EDTA" from Roche), and centrifuged at 16,000 g for 20 min at 4°C. Inclusion bodies were prepared by resuspending the cell pellet in Bug Buster reagent (28 ml) supplemented with lysozyme (200 μg/ml), incubated at room temperature for 5 min with proteases inhibitors, and centrifuged at 16,000 g for 20 min at 4°C. The cell pellet was further washed with 120 ml of 1:10 diluted Bug Buster reagent with protease-inhibitors (in deionized water), vortexed for 1 min, and centrifuged at 16,000 g for 20 min at 4°C. This step was repeated a second time with only 40 ml of 1:10 diluted Bug Buster reagent. Inclusion bodies were solubilized in 20 ml of 20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, 6 M guanidine hydrocloride (GdnHCl), and 5 mM β-mercaptoethanol (βME). Finally, the preparation of the inclusion bodies was clarified by centrifugation at 20,000 g for 20 min (4°C) and the supernatant was filtered through a 0.2 μm syringe filter.

The His-tagged protein solubilised from inclusion bodies was purified on a self-packed 2 ml immobilized-metal affinity column charged with nickel (Novagen) and equilibrated in 20 mM sodium phosphate-pH 7.4, 0.5 M NaCl, 6 M GdnHCl, and 5 mM βME (equilibration buffer). The column was operated under gravity flow at 4°C. The preparation of inclusion bodies was loaded in 10 ml aliquots, and the column was washed with 20 ml of equilibration buffer applied in 4 batches of 5 ml. Elution was achieved by an imidazole concentration step-gradient. Each step was 5 ml of equilibration buffer supplemented with increasing concentration of imidazole, as follows: 10, 20, 30, 50, 80, 90, 100, 500 mM and 1 M. Most of the protein was recovered in the 30-90 mM imidazole fractions (Figure 3.17). These fractions were enriched in FTM2 when analysed by SDS-PAGE, although all showed contamination, with smaller fragments and a higher molecular weight host protein (~30 kDa). Further analysis by Western blotting using anti-His6 antibodies suggested that some of the smaller fragments came from proteolytic degradation of the FTM2 at its C-terminus.

Interestingly, the amount of recombinant protein recovered from large-scale growths was far smaller than expected from the amount observed in small-scale expression tests. The use of denaturing conditions, along with the use of proteases inhibitors throughout purification (although EDTA was not employed), led to speculation that degradation might have occurred during cell

growth. Thus, this construct was not pursed, and use of the thioredoxin solubility tag was investigated.

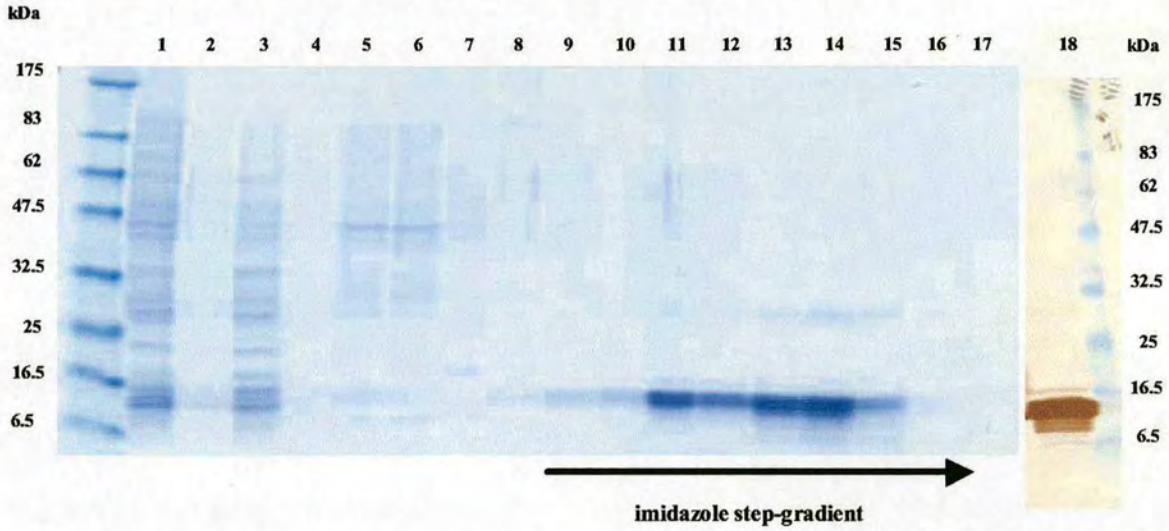



**Figure 3.17: FTM2-pET15b purification by immobilised-$Ni^{2+}$ affinity chromatography**

Lane 1: inclusion bodies extract (10 µl from a 10 ml fraction). Lane 2: unbound fraction (10 µl from a 10 ml fraction). Lane 3: as lane 1. Lane 4: as lane 2. Lane 5-8: buffer wash fraction, 50 µl were TCA precipitated from a 5 ml fraction. Lane 9-17: imidazole step gradient in buffer-10, 20, 30, 50, 80, 90, 100, 500 mM and 1 M imidazole-each lane, each lane is 50 µl TCA precipitated from a 5 ml fraction. Gel is 10-20% SDS Tris-HCl (reducing conditions), and visualized using the Coomassie blue method. Lane 18: Western blot of lane 14 using anti-His6 antibodies (membrane staining as described in section 3.2.3.3).

### 3.3.3.2 Fibulin-type module from fibulin-2 expressed as fusion proteins to enhance solubility

#### 3.3.3.2.1 pET32a(+) construct

The DNA insert encoding for FTM2 with a 5' *NcoI* and a 3' *XhoI* restriction site was excised from the YTH vector (pACT2-Clontech) provided by Dr. Julia White. The insert was further ligated in the pET32a(+) vector (Novagen; Figure 3.18) previously digested by the same restriction enzymes. The resulting construct was named FTM2-pET32a(+).

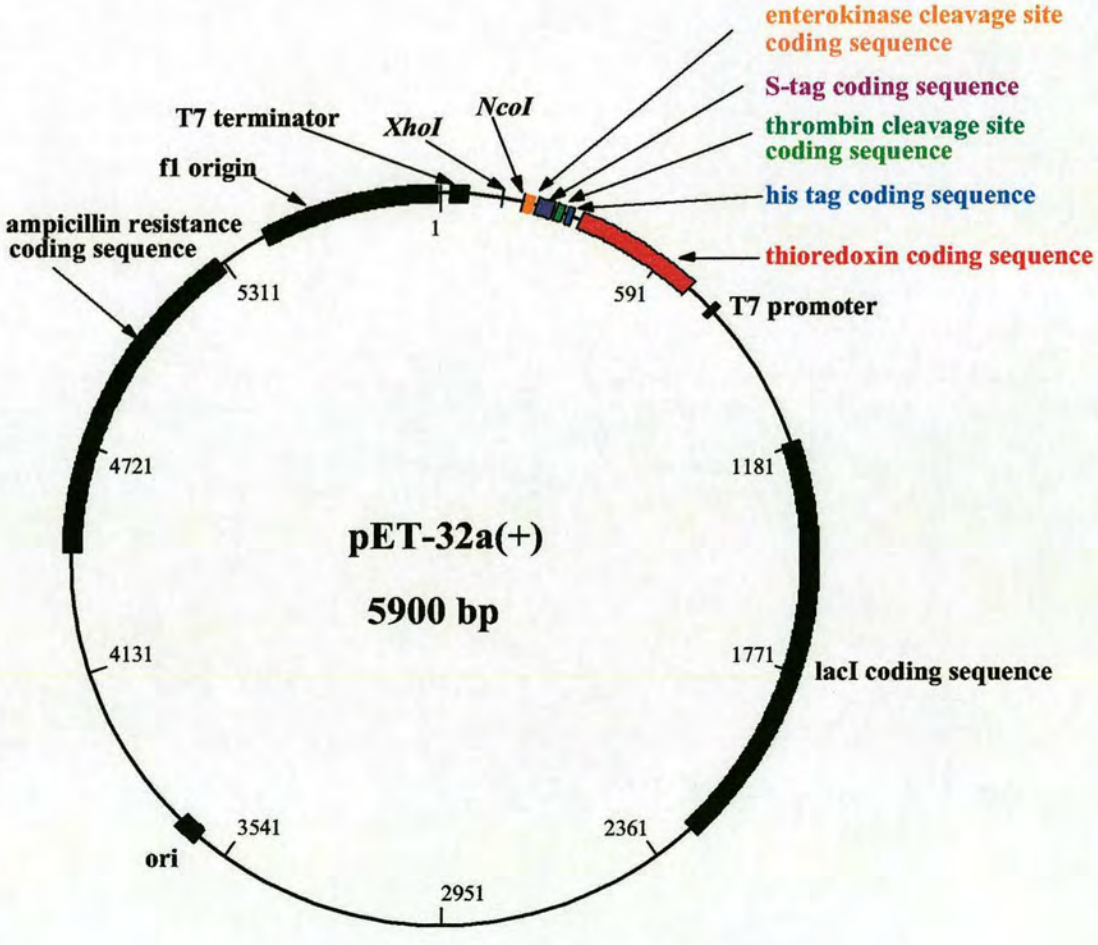



**Figure 3.18: pET32a(+) vector from Novagen**

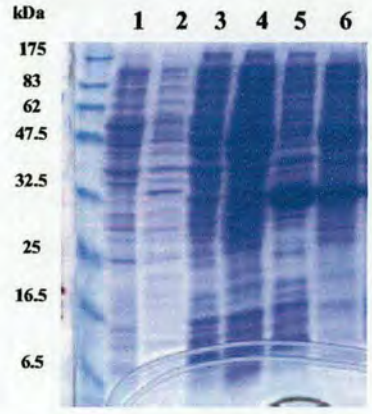



**Figure 3.19: FTM2-pET32a induction test**

Lane 1: *E. coli* BL21(DE3) -control-insoluble fraction. Lane 2: same as lane 1-soluble fraction. Lane 3: *E. coli* BL21(DE3)-transformed with the FTM2-pET32a(+) non induced- insoluble fraction. Lane 4: same as lane 3 but induced- insoluble fraction. Lane 5: same as lane 3 but non induced- soluble fraction. Lane 6: same as lane 3 but induced- soluble fraction. Gel is 10-20% SDS Tris-HCl (reducing conditions), and visualized using the Coomassie blue method. Test and sample load according to section 3.2.2.1

Induction tests and large-scale growth were as described for the pET15b construct. When fused to thioredoxin (Tdx), the FTM2 is expressed as a soluble protein, and to a high level (Figure 3.19). The FTM2 fused to thioredoxin (FTM2-pET32a(+) is ~30 kDa) was purified by immobilized $Ni^{2+}$-affinity chromatography. Several large-scale preparations were made with 4 h of induction (1 mM IPTG), and cell pellets were between 3-4 g. Soluble extracts were prepared by use of the Bug Buster reagent as described previously. Elution was achieved by imidazole step gradient: 10, 20, 30, 40, 50, 80, 90, 100, 150, 200, 300, 500 mM and 1 M imidazole in equilibration buffer, as in purification of the pET15b construct. Most of the fusion protein was recovered in the 50-90 mM imidazole fractions, and gave a single band by SDS-PAGE analysis (Figure 3.20-a).

These fractions were pooled and desalted against different buffers by dialysis or ultrafiltration (pressure cell). Desalting against enterokinase cleavage buffer (20 mM Tris-HCl - pH 7.4, 50 mM NaCl, 2 mM $CaCl_2$) and thrombin buffer (20 mM Tris-HCl - pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$) is necessary to release the FTM2 from Tdx. In both cases, however, desalting against these buffers resulted in protein precipitation, with only a small proportion of the protein preparation recoverable.

Cleavage with enterokinase (25 units in 5 ml-24 h at 20°C) led to further protein precipitation during incubation, and to non-specific digestion products. Thrombin cleavage (12 units in 5 ml - 24 h at 20°C) led to the expected products: the total fusion protein has a predicted molecular weight of 30.8 kDa, which after cleavage lead to two distinct fragments of ~14 kDa and ~16.5 kDa. The smallest corresponds to the Tdx with a His6 tag, while the largest corresponds to the FTM2 with an extra streptavidin-affinity tag (S-tag) and an enterokinase cleavage site (Figure 3.20-b).

After cleavage, the mixture was diluted in the immobilized-$Ni^{2+}$ affinity chromatography equilibration buffer (20 ml) and re-loaded on a $Ni^{2+}$-column that had been pre-equilibrated in the same buffer. The column was then washed with 5 ml of buffer, and elution of the bound protein was achieved with a five-step imidazole gradient (10, 20, 30, 50, 100 mM - each time 5 ml). Fractions were analysed by SDS-PAGE. The FTM2 was found in the unbound and wash fractions, while Tdx was found mainly in the 30-100 mM fractions (Figure 3.20-c). Both unbound and bound fractions eluting at 50 mM imidazole were analysed by Western blotting with anti-His6 antibodies (Figure 3.20-d)

Overall only a small quantity of FTM2 was recovered, and subsequent investigations such as refolding or biophysical characterisation were impossible. Moreover, cleavage of the fusion protein with thrombin results in the FTM2 still fused to an S-tag, an affinity tag that would certainly interfere with antibody-production and possibly with refolding.

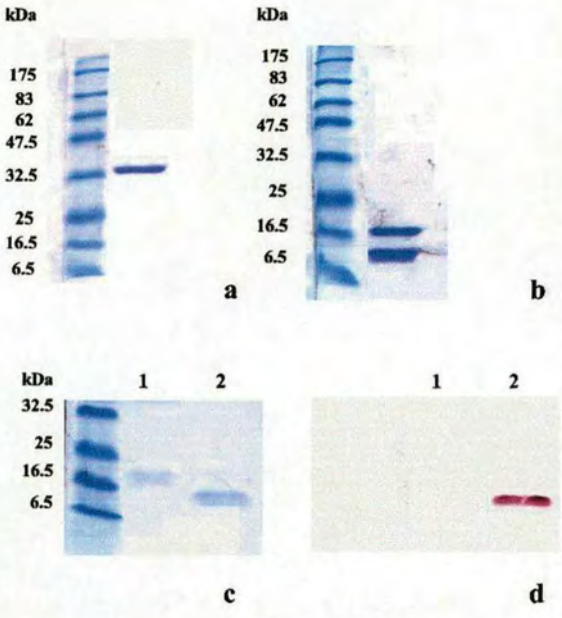



**Figure 3.20: Purification of the FTM2-pET32a(+) construct**

**a)** SDS-15% PAGE analysis under reducing conditions of the FTM2-pET32a(+) construct (30.8 kDa) before thrombin cleavage - 100 µl were TCA precipitated from a 5 ml fraction. Gel visualized using the Coomassie blue method.

**b)** SDS-15% PAGE analysis under reducing conditions of the FTM2 (16.7 kDa with tags) and Tdx (14.1 kDa) free products from thrombin cleavage - 300 µl were TCA precipitated from a 5 ml fraction. Gel visualized using the Coomassie blue method.

**c)** SDS-15% PAGE analysis under reducing conditions of the cleavage products after reload of (b) on nickel-affinity resin. Lane 1 is the unbound fraction, and lane 2 is the bound fraction eluted with 50 mM imidazole; each time, 1 ml was TCA precipitated from a 20 ml fraction (lane 1) and from a 5 ml fraction (lane 2). Gel visualized using the Coomassie blue method.

**d)** Western blot of (c) using anti-His6 antibodies (membrane staining as described in section 3.2.3.3).

#### 3.3.3.2.2 pET1532 construct

Complete removal of affinity tags was thought to be achieved by enterokinase cleavage in the pET32 construct, as such cleavage was found to be non-specific (above), a strategy based on compatible restriction sites between the FTM-pET15b and pET32a(+) vectors (Figure 3.15 and 3.18) was used to introduce a thrombin site prior to the FTM2 sequence (Figure 3.16). This approach was developed to avoid further PCR engineering of the FTM2 DNA coding sequence; also the presence of an extra His6 tag was found to improve one-step purification of the fusion protein.

The DNA encoding for the FTM2 fused (at its N-terminus) to a thrombin cleavage sequence and a His6 tag was obtained by digesting the FTM2-pET5b plasmid using *NcoI* and *XhoI* restriction sites. This insert was then ligated in the pET32a(+) vector previously digested with same restriction enzymes. The resulting construct was named FTM2-pET1532.

Large-scale preparations and one step purification of the fusion protein by immobilised-nickel affinity chromatography were as described for the FTM2-pET32a(+) construct. Fusion protein eluted at 100-250 mM imidazole, but mainly between 150 and 200 mM. This increase in binding illustrates the utility of the second His6 tag. The fusion protein was processed as described above. Despite extensive protein precipitation during desalting and concentration; some protein was recovered (Figure 3.21). Use of biotinylated thrombin allowed its removal after cleavage by streptavidin agarose (Thrombin cleavage/capture kit, Novagen). Fractions seen in Lane 3, 4 and 5 in Figure 3.21 were pooled, concentrated and submitted for antibody-production to Dr. Jeffrey McIlhinney.

Diverse approaches were tested to minimise or recover some of the precipitate. One approach was to concentrate the fusion protein against a weak denaturing buffer to prevent protein precipitation. Fractions from $Ni^{2+}$-affinity chromatography were concentrated and desalted with 10 mM Tris-HCl, pH 8, 3 M urea, to a volume of 4.5 ml. The protein was then diluted once with 10 mM Tris-HCl, pH 8, to give a final concentration of 1.5 M urea. One ml of 10x thrombin cleavage buffer was added to the 9 ml, and cleavage was performed as described previously but with 50 thrombin units. After cleavage the protein was diluted to 20 ml of equilibration buffer containing 6 M urea and 10 mM reduced glutathione (GSH), and $Ni^{2+}$-affinity chromatography was performed under the same denaturing and reducing conditions (Figure 3.22).

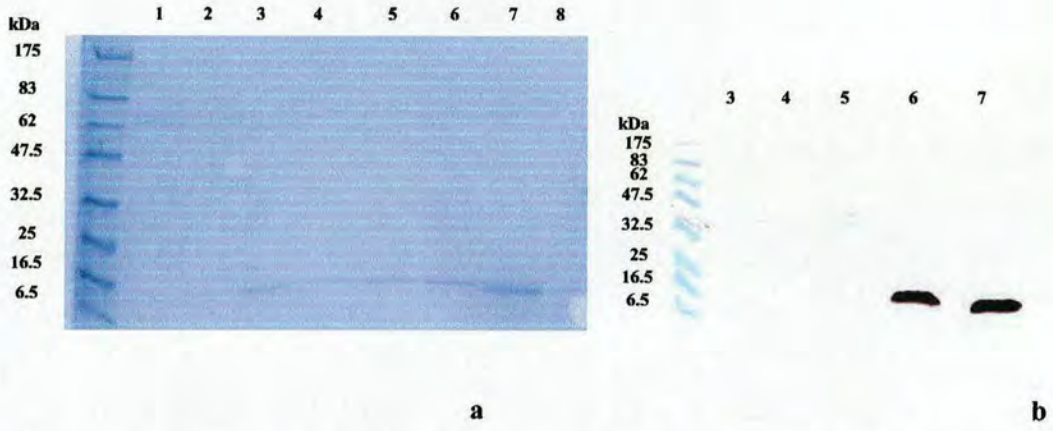



**Figure 3.21: Purification of the FTM2-pET1532 construct under native conditions**

**a)** Lane 1: unbound fraction - 200 µl were TCA precipitated from a 10 ml fraction. Lane 2: buffer wash fraction - 200 µl were TCA precipitated from a 5 ml fraction. Lane 3-8: imidazole step gradient in buffer-10, 20, 30, 50, 100, 500 mM imidazole- each lane is 200 µl TCA precipitated from a 5 ml fraction. Gel is 10-20% SDS Tris-HCl (reducing conditions), and visualized using the Coomassie blue method.

**b)** Western blot of lanes 3-7 from (a) using anti-His6 antibodies (membrane staining as described in section 3.2.3.3).

Surprisingly, some uncleaved fusion protein was found in the unbound fraction, wash and early gradient fractions along with the FTM2; while Tdx bound to the column and eluted later in the gradient as shown by Western blotting (anti-His6 antibodies; Figure 3.22). One explanation is that thrombin cleavage under denaturing conditions is less efficient, thus leading to a significant amount of uncleaved fusion protein, which in this preparation did not totally bind onto the $Ni^{2+}$ affinity resin. However, it will be possible to isolate 2-3 mg of FTM2 by adding a gel-filtration step to this last approach. This could then enable antibody-production.

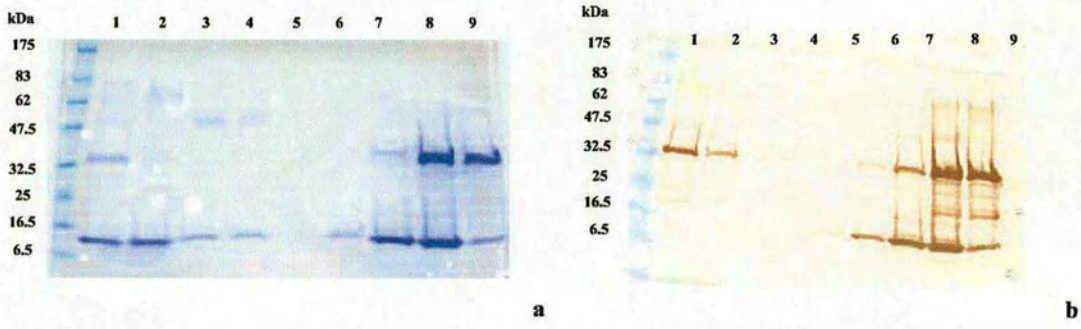



**Figure 3.22: Purification of FTM2-pET1532 construct under denaturing conditions**

**a)** Lane 1: unbound fraction -100 µl were TCA precipitated from a 10 ml fraction. Lane 2-9: imidazole step gradient in buffer-20, 40, 60, 80, 100, 150, 200, 500 mM imidazole- 6 M urea, 10 mM GSH- each lane is 50µl TCA precipitated from a one ml fraction. Gel is 10-20% SDS Tris-HCl (reducing conditions), and visualized using the Coomassie blue method.

**b)** Western blot of (a) using anti-His6 antibodies (membrane staining as described in section 3.2.3.3).

# CHAPTER FOUR: BIOPHYSICAL STUDIES OF RECOMBINANT PROTEINS

## 4.1 Nuclear Magnetic Resonance spectroscopy studies

The $GABA_B$ R1a CP fragments investigated in the current study were presented in chapter three, and their order of appearance was related chronologically to the iterative process that took place between the protein production phase of this work and its goal -NMR spectroscopy. This chapter presents the NMR characterization of these fragments along with the hypotheses regarding the folding of the CP modules formulated during the course of this study.

$^{1}H$-$^{15}N$ HSQC spectra provide a powerful tool with which to investigate protein folding. In such spectra, resonances arise from correlations between nitrogen nuclei and their attached protons. Therefore these spectra constitute "protein fingerprints" with one cross-peak for each non-proline residue in the molecule. Good chemical shift dispersion is characteristic of well-folded proteins. Figure 4.1 shows the $^{1}H$-$^{15}N$ HSQC spectrum of TR-CP12. The spectrum displays 114 cross-peaks of different intensities (count exclude side chain $NH_2$ from Gln or Asn residues, and $NH\varepsilon_1$ from Trp residues). TR-CP12 contains 136 residues plus four extra residues, which are cloning artifacts. TR-CP1 is predicted to contain 72 residues, while CP2 is predicted to consist of 61 amino acids, and there is a three-residue linker between the two modules; there are also 9 proline residues. Thus, the number of cross-peaks found in the $^{1}H$-$^{15}N$ HSQC spectrum of TR-CP12 is close to the number of cross-peaks that one would expect for two folded modules. However in 2D-$^{1}H$ and 3D-$^{15}N$ edited TOCSY experiments only few cross-peaks were detected (Figure 4.2).

Missing cross-peaks in protein TOCSY spectra can arise from severe line broadening and this may be due to several factors: aggregation leads to multimers that have shorter $T_2$ values than the monomeric form and in general result in line broadening; conformational fluctuations on an intermediate time scale results in line broadening; amides in intermediate exchange with water can also be a cause of line broadening. Aggregation is often a problem in protein NMR because samples must be at 0.5 – 1.5 mM concentration (in standard probes) to achieve adequate signal-to-noise ratios. The correct choice of temperature, pH, buffer type, and ionic strength may help to reduce some of the factors that contribute to line broadening.

Several samples of the TR-CP12 were produced and tested in various conditions of temperature (12-37°C), pH (Figure 4.3), and buffers (Tris buffer, acetate buffer, phosphate buffer).

Two types of detergents, 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) (5-25 mM) and 8,octyl-β-D-gluco-pyranoside (5-20 mM) were used. Additions of glycerol (5-10%) or increasing concentrations of sodium chloride (up to 500 mM), and addition of GABA (up to 6 mM), or $Ca^{2+}$ (up to 20 mM) were also tested. In addition to inspection of TOCSY experiments, one-dimensional $^{15}N$ $T_2$ measurements [164] were used to monitor the effects of these conditions. $^{15}N$ $T_2$ values were estimated from 1D $^{1}H$-$^{15}N$ HSQC spectra recorded with different relaxation delays, and enabled an estimate of the mean $^{15}N$ $T_2$ value for the ensemble of residues (Figure 4.4). Best conditions for the TR-CP12 were established as 20 mM sodium actetate, pH 4, at 37°C, which yielded a mean $^{15}N$ $T_2$ of 78 ms (0.6 mM). This value is shorter than the one observed for another CP module pair; the $^{15}N$ $T_2$ of VCP~23 were approximately 100-110 ms. On the other hand the TR-CP12 mean $T_2$ value was not much lower than that of the CR1~15-16 pair, which exhibited values of around 85-90 ms. A doubling of the protein concentration (1.2 mM) resulted in a 10 ms decrease of the mean $^{15}N$ $T_2$ value, suggesting that the protein is prone to aggregation (not shown). Nevertheless, the quality of the 2D-$^{1}H$ and 3D-$^{15}N$ edited TOCSY experiments of TR-CP12 did not improve and it was concluded that beside a possible aggregation problem, a protein folding issue could be responsible for the shortcomings of the TOCSY spectra.

N-glycosylation can have a profound effect on protein folding. Some proteins have been found to fold properly only when correct glycans were attached onto the polypeptide chain [165, 166]. An NMR study of the adhesion domain of the human CD2 [167], a cell surface glycoprotein found lymphocytes, revealed that interactions between specific residues and the N-glycan were essential for proper folding of the molecule. Strikingly, only the first two GlcNac units from the sugar moiety were involved in these contacts. In our laboratory, Dr Joanne O'Leary reported that mutation of the N-glycosylation consensus sequence in the membrane cofactor protein CP module 1 altered its folding (Dr Joanne O'Leary- PhD thesis). Consequently a sample of the glycoform of TR-CP12 was prepared (see chapter three) and investigated by NMR. Some differences can be observed from a direct comparison of the two $^{1}H$-$^{15}N$ HSQC spectra (Figure 4.6). Six to eight extra cross-peaks can be observed, some cross-peaks observed in the non-glycoform are shifted or became more intense, and two other cross-peaks disappeared (in the region ~117 ppm in $^{15}N$ and~7.5 ppm in $^{1}H$). These differences are, however, difficult to interpret. Some cross-peaks might be shifted simply because of the presence of the N-glycan rather than because of a change in folding, also it worth noting that this sample is slightly more concentrated. However the mean $^{15}N$ $T_2$ value of the glycosylated form was similar to the non-glycosylated form of TR-CP12 (83 ms) and the quality of TOCSY experiments did not significally improved.

The presence of the seven-residue extension at the beginning of CP1 might be of importance in terms of protein folding. In the case of the immunoglobulin-type domains of titin, a large multiple modular muscle protein, the amino-terminal extension (six residues) was found to have a profound effect on protein stability [168, 169]. The $^{1}H$-$^{15}N$ HSQC spectrum of the $GABA_B$ R1a~"full-length and mutated to remove N-glycosylation site 1" CP12 (FLM-CP12) construct did not, however, support this hypothesis; only a few extra sharp cross-peaks can be observed when compared to the truncated version (Figure 4.6), and no significant improvement in the quality of the TOCSY experiments was observed from this construct. Unfolded residues are characterised by sharp resonances, and from this observation it can be speculated that the N-terminal extension is unfolded.

At this stage of the study, the choice could have been to attempt assignment of TR-CP12 using triple resonance experiments. An important motivation for working with the module pair rather than the single modules was to elucidate details of the intermodular orientation. This depends upon a full assignment of residues near the junction between modules, and the observation of critical intermodular NOEs. However, on the basis of the overall quality of the HSQC, and the lack of signals in 2D-$^{1}H$ and 3D-$^{15}N$ TOCSY spectra, it was judged unlikely that this sample would yield sufficient NMR data for a useful structure determination. It was therefore decided that investigation of individual modules should be the next step, with the possibility of returning to the double module at a future date. Single module constructs were therefore produced to investigate the folding of individual modules (Figure 4.7).

The $^{1}H$-$^{15}N$ HSQC spectrum of the CP2 (Figure 4.7) exhibits a good chemical shift dispersion characteristic of a well-folded domain. Surprisingly, there are more cross-peaks in the spectrum than expected – 86 peaks (count excludes the region containing side-chain resonances of Asn, Gln, and Trp) can be found instead of the 64 expected. It was concluded that CP2 has more than one conformation in solution, with the extra cross-peaks arising from the existence of additional slowly- or non-exchanging conformation(s) possibly in a localized region of the polypeptide chain, later attributed to a *cis/trans* proline isomerisation (see chapter five).

Subsequently, the $^{1}H$-$^{15}N$ HSQC spectrum of CP1 expressed as an isolated module was found to have fewer dispersed cross-peaks than expected, with little chemical shift dispersion and a significant number of random-coil position peaks, indicating that it does not have a clearly defined structure (Figure 4.7). It may be largely unfolded, or may exist as multiple conformations in intermediate exchange. Further investigations of TR-CP1 included changing temperature (8-37°C) and pH titration (3.5-8) were undertaken without success (Figure 4.8). Also investigated were

comparisons of TR-CP1 with $GABA_B$ R1a~"wild-type" (WT-CP1) (Figure 4.9), and of the fully glycosylated WT-CP1 with its deglycosylated form (Figure 4.10). Finally the WT-CP12 glycosylated (Figure 4.11) and the WT-CP12 non-glycosylated were also compared (not shown). None of these constructs yielded NMR spectra that showed any significant improvements, although some resonances seem to get sharper in the "wild-type" versions when compare to the "truncated" ones.

Superposition of the $^{1}H$-$^{15}N$ HSQC spectra, and of the $^{1}H$ 2D-NOESY spectra, collected on CP12 and CP2 (Figure 4.12, 4.14, and 4.15), show that some of the well-resolved cross-peaks corresponding to CP2 are shifted in the context of the pair, and therefore some interaction might exist between the two modules. Interestingly, the ratio of *trans* to *cis* conformer population of CP2 in the context of the pair appears to differ from isolated CP2 (Figure 4.12). This could be assess through well resolved cross-peaks from CP2 inside the module pair, and will be the focus of future work.

CP1 appears to be more structured in the context of the module pair than on its own (Figure 4.14, and 4.15). Interestingly, ~11 resonances in the TR-CP1 $^{1}H$-$^{15}N$ HSQC spectrum (~108-113 ppm in $^{15}N$, and ~ 8.25-8.7 ppm in $^{1}H$) have shifts that are characteristic of the glycine/theronine residues in unstructured peptides [170], and therefore could account for the seven glycine and four threonine residues found in its primary sequence (Figure 4.7). The $^{1}H$-$^{15}N$ HSQC spectrum of CP1 has similarities with the $^{1}H$-$^{15}N$ HSQC spectrum of the human α-lactalbumin (α-LA) molten globule formed at pH 2, where only three of the 122 backbone resonances can be observed due to extensive conformational fluctuations throughout the protein on a millisecond time scale [171]. Upon full denaturation of the α-LA molten globule, additional resonances appeared. Because the resonances of the denatured α-LA had been characterised, these additional peaks were assigned, and it was possible to identify the region of the α-LA molten globule that unfolds. Interestingly, some cross-peaks observed in the denatured α-LA never appeared in the denatured α-LA molten globule. These peaks were thought to exhibit extensive broadening, and it was speculated that they may correspond to regions of the molecule that remain in a partly-folded, collapsed state, with some residual structure. It is difficult to draw conclusions about the extent to which CP1 is folded or in the molten globule form. Some resonances have chemical shifts characteristic of unstructured peptides, others are consistent with folded structure. On the other hand, the broadening of some resonances evident in the $^{1}H$-$^{15}N$ HSQC spectrum of the CP1 (Figure 4.7) seems consistent with slow fluctuations of the polypeptide chain, as seen in molten globules, which arise from residual structure restricting the conformational freedom of the polypeptide chain.

Studies on protein folding suggest that molten globules retain native-like secondary structure [172] and correct folding topology even in the absence of cooperative interactions between side-chains [173]. The classic molten globule, therefore, lacks the side-chain interactions that characterize the tertiary structure of the native protein [174]. A few examples in the literature, however, emphasise the persistence of tertiary structure in some molten globules [173, 175], which may be either native-like or not, and the term "highly-ordered" [174] molten globule has been used to describe such intermediates between native and molten globule states. In $^{1}$H 2D-NOESY experiments, resonances arise from dipole-dipole interactions between protons through space, and in the range of 5 Å. Cross-peaks that correlate protons from different side-chains are therefore indicative of protein folding. Analysis of the $^{1}$H 2D-NOESY spectrum for the TR-CP1 suggests the retention of tertiary interactions as some NOEs between the aromatic and the aliphatic side-chains can be observed (see Figure 4.13). Evidence for proper disulphide bond formation in CP1 was presented in chapter three; such an observation suggests the retention of some tertiary structure.

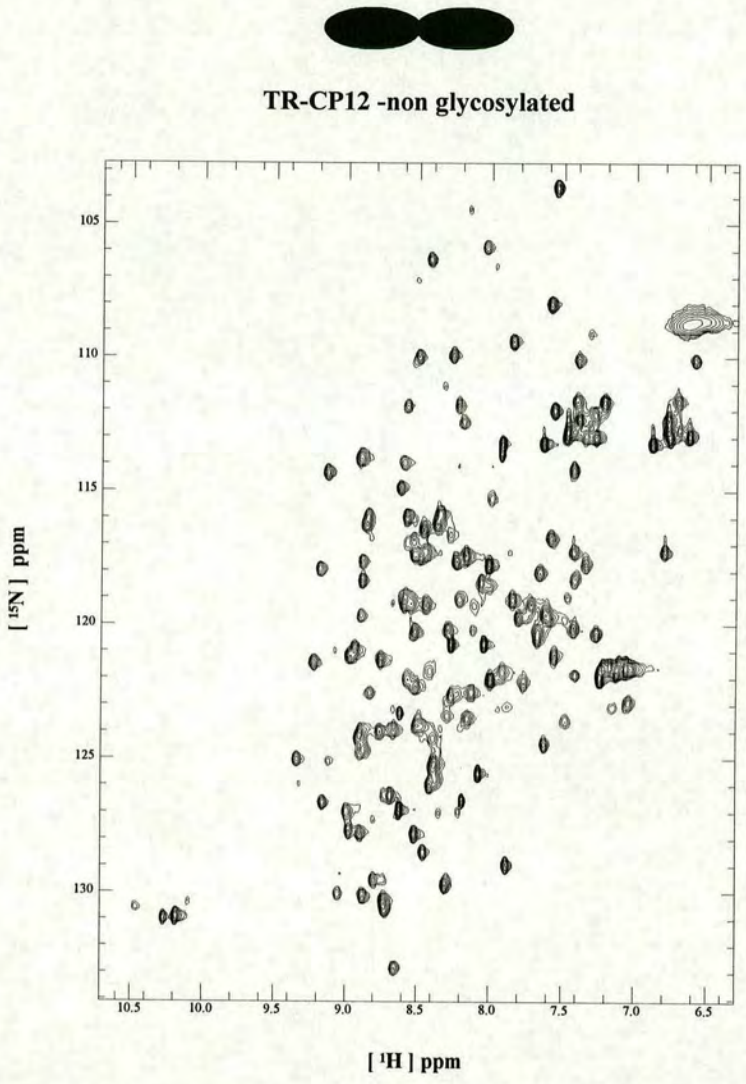



**Figure 4.1: $^{1}$H-$^{15}$N HSQC-spectra of TR-CP12 -non glycosylated**

Sample was 0.6 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$.

Spectrum was acquired at 600 MHz, and 37°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2212.6 Hz in the $^{15}$N dimension. Spectrum was recorded with 10 scans, 1024 complex points in the $^{1}$H dimension, and 160 complex points in the $^{15}$N dimension.

After processing, data matrix was 1024 points in $^{1}$H, and 256 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

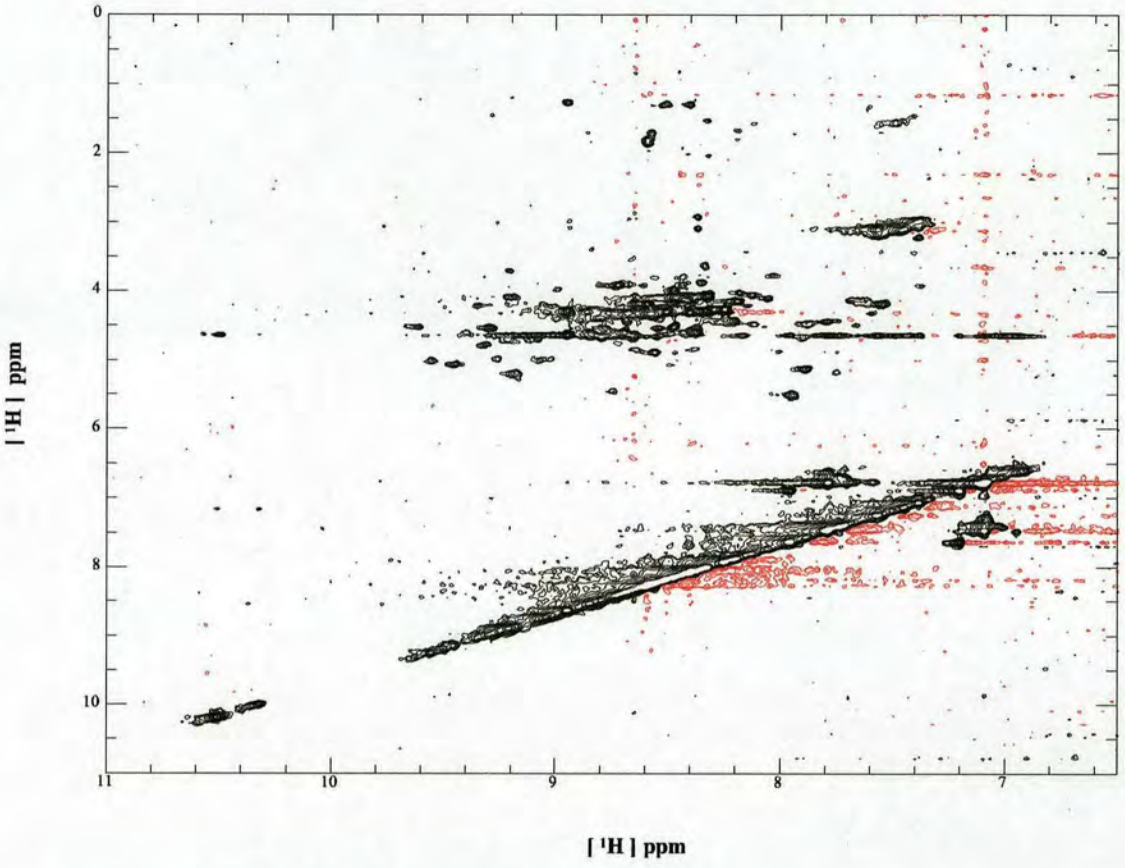



**Figure 4.2: 3D $^{1}H$-$^{15}N$ edited-TOCSY-spectrum of TR-CP12 recorded with only one increment in the $^{15}N$ dimension**

Sample was 0.6 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 37°C. Mixing time was 0.035 s.

Sweep width is 5000 Hz in the directly detected $^{1}H$ dimension, 7400 Hz in the indirectly detected $^{1}H$ dimension. Spectrum was recorded with 40 scans, 1024 complex points in the acquisition dimension, and 256 complex points in the indirectly detected $^{1}H$ dimension. After processing, data matrix was 1024 points in the acquisition dimension, 512 points in the indirectly $^{1}H$ dimension. A sine function squared (with an angle of 90°) was applied in all dimensions. Positive contours levels are colored black and negative contours levels are colored red.

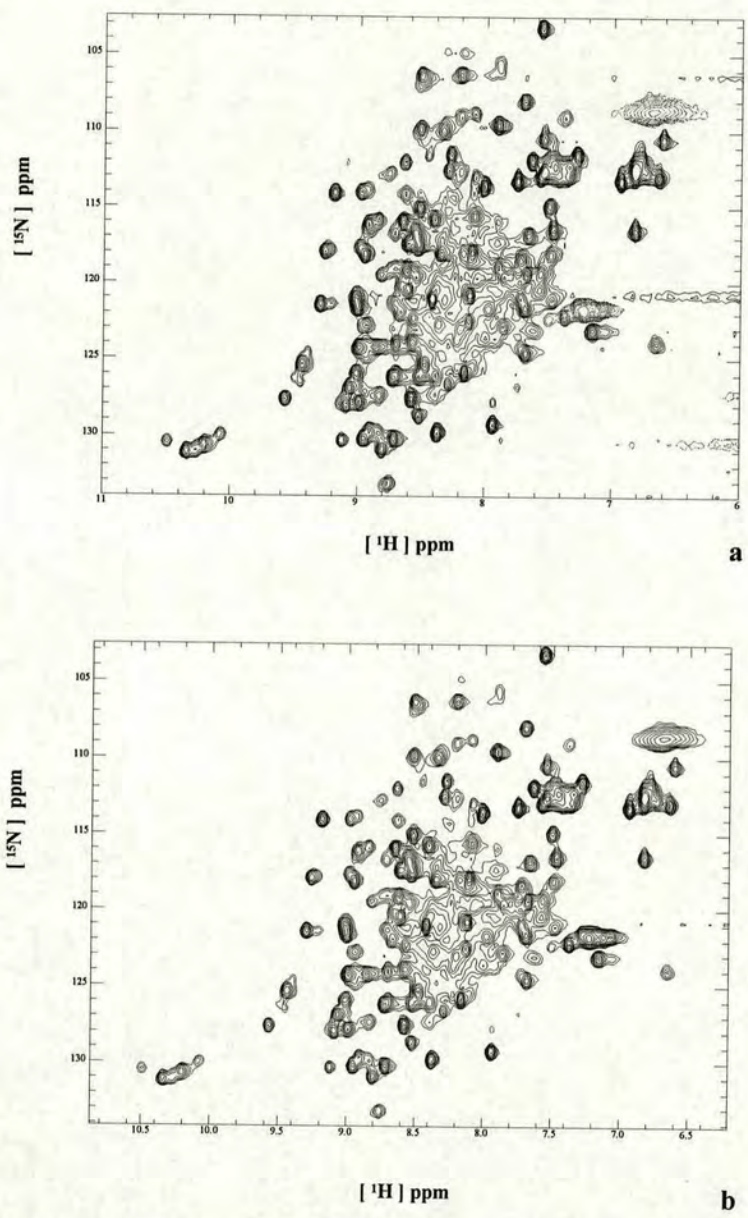



**Figure 4.3: pH titration of TR-CP12**

**a) TR-CP12 -non glycosylated at pH 4**

Sample was 1.1 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 37°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2000 Hz in the $^{15}$N dimension. Spectrum was recorded with 16 scans, 1024 complex points in the $^{1}$H dimension, and 96 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 128 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) TR-CP12 -non glycosylated at pH 5.5**

Same experimental conditions as in (a), pH was adjusted with NaOH.

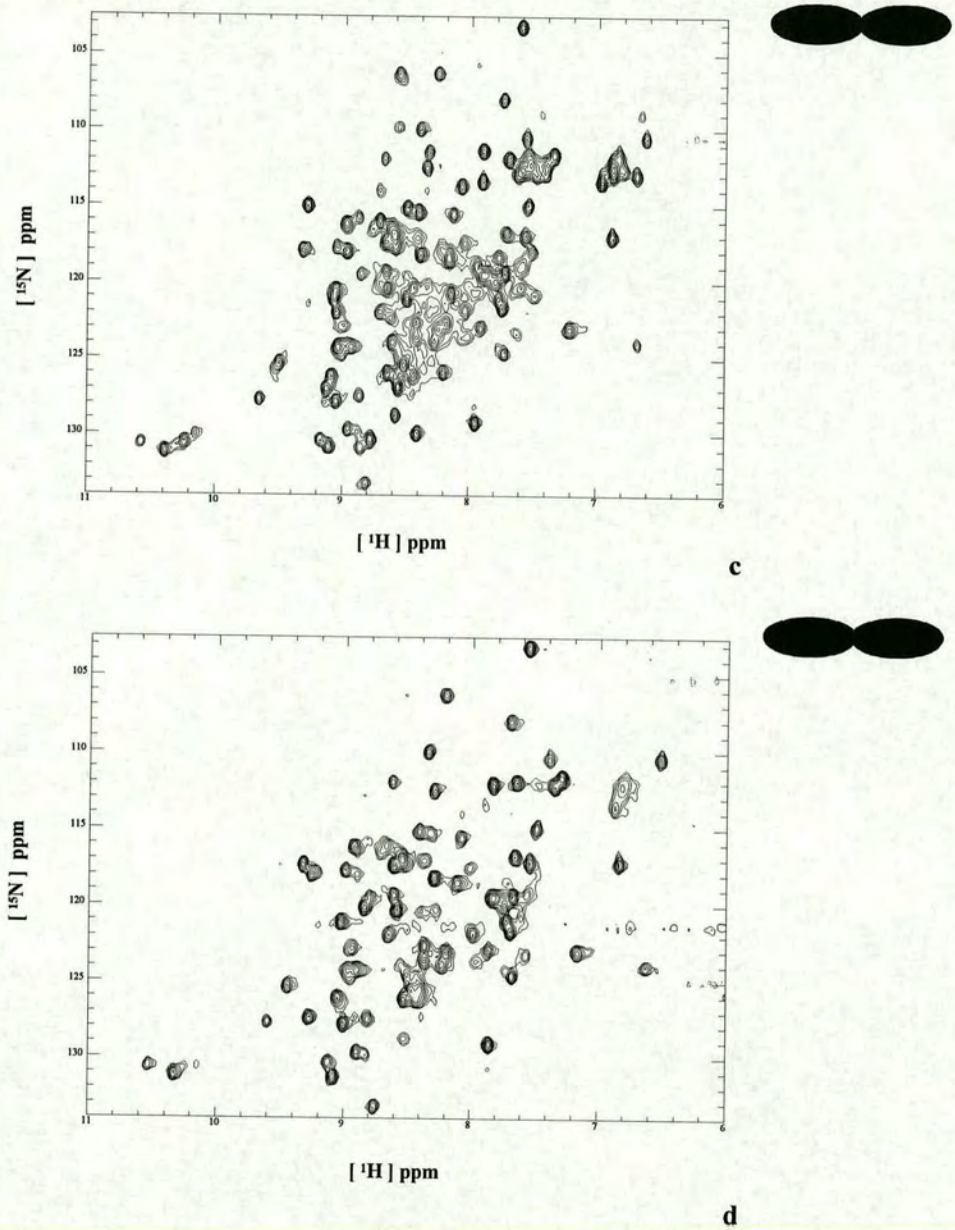



**Figure 4.3 continued: pH titration of TR-CP12**

**c) TR-CP12 -non glycosylated at pH 7.3**

Same experimental conditions as in (a), pH was adjusted with NaOH.

**d) TR-CP12 -non glycosylated at pH 8.5**

Same experimental conditions as in (a), pH was adjusted with NaOH.

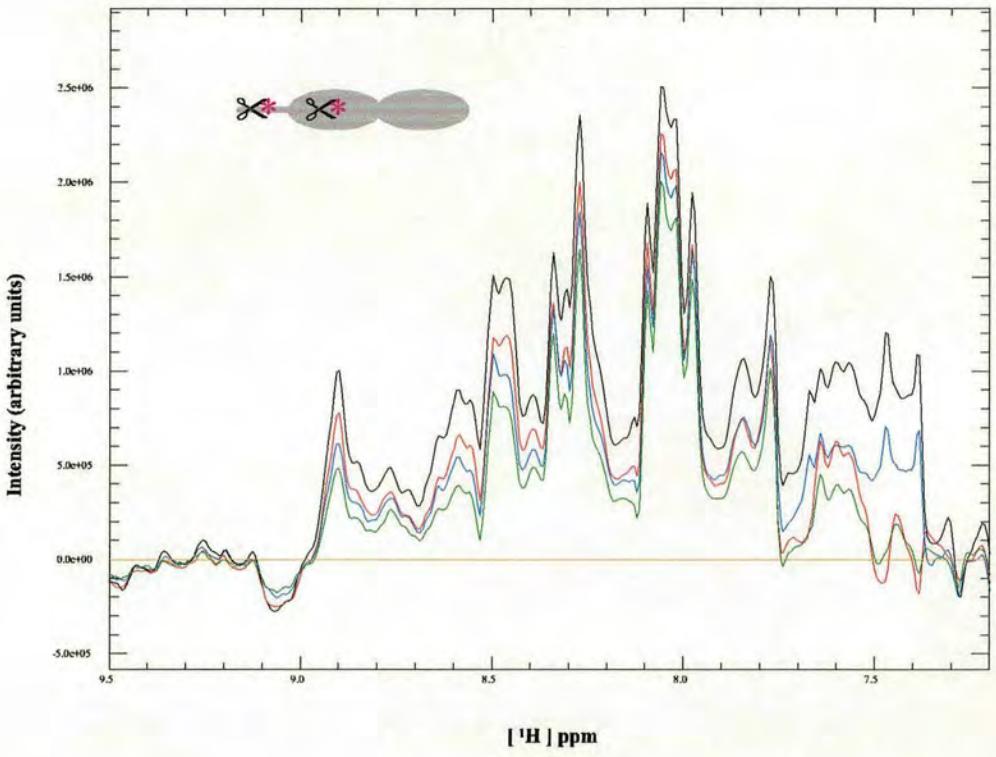



**Figure 4.4: $^{15}N$ $T_2$ measurement for WT-CP12-deglycosylated**

Sample was 0.4 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 37°C.

$^{15}N$ $T_2$ -1D trace from $^{1}H$-$^{15}N$ HSQC recorded with different relaxation delays: 16 ms (black), 32 ms (red), 48 ms (blue), and 64 ms (green).

Spectra were processed in the AZARA software package, and the relative intensity ratio between two spectra was used to calculate $^{15}N$ $T_2$ from the following: $^{15}N$ $T_2$ = (time delay spectrum #2 – time delay spectrum #1) / ln (intensity ratio). Sweep width is 5000 Hz in the $^{1}H$ dimension (1024 complex points), and 2212.6 Hz in $^{15}N$. A sine function squared (with an angle of 90°) was applied. 128 scans per delay were collected.

Average $^{15}N$ $T_2$ value is 80 ms in the 7.8 to 9 ppm range.

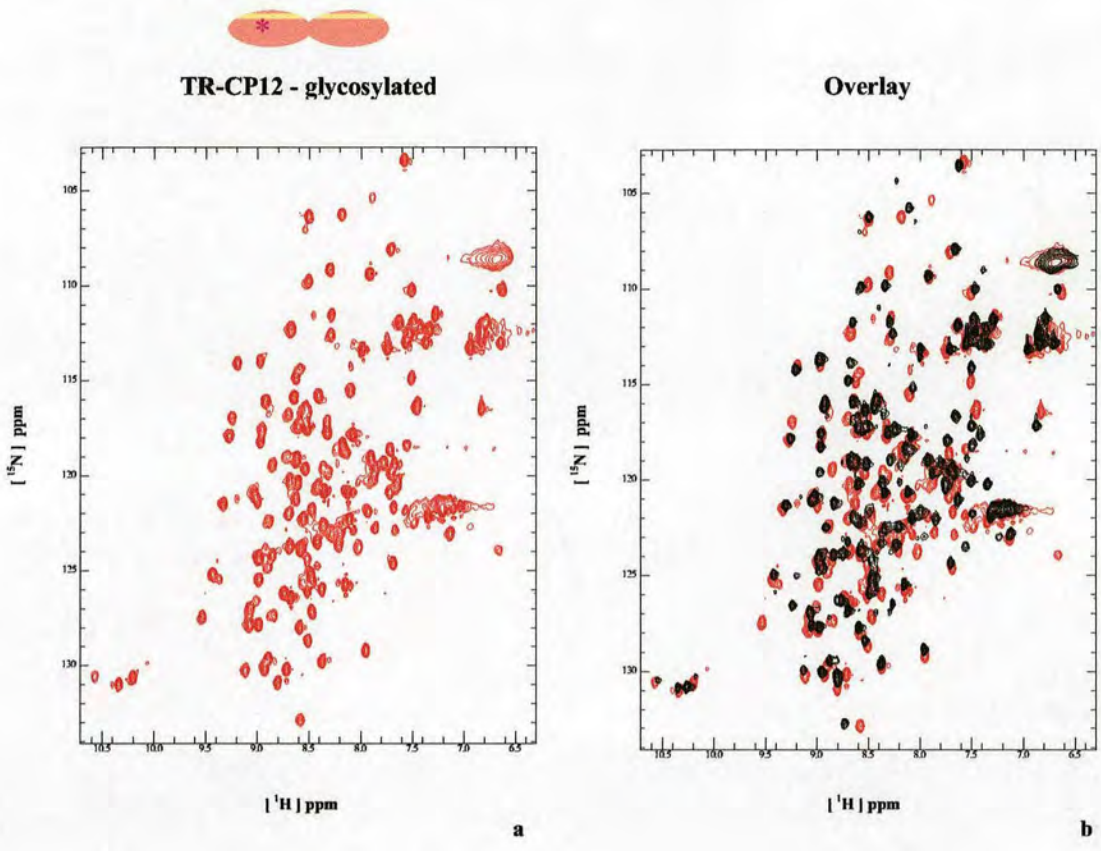



**Figure 4.5: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) TR-CP12 – glycosylated**

Sample was 0.8 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$.

Spectrum was acquired at 600 MHz, and 37°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2212.6 Hz in the $^{15}$N dimension. Spectrum was recorded with 10 scans, 1024 complex points in the $^{1}$H dimension, and 160 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 256 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) TR-CP12 -non glycosylated (black) and TR-CP12 – glycosylated (red) overlaid**

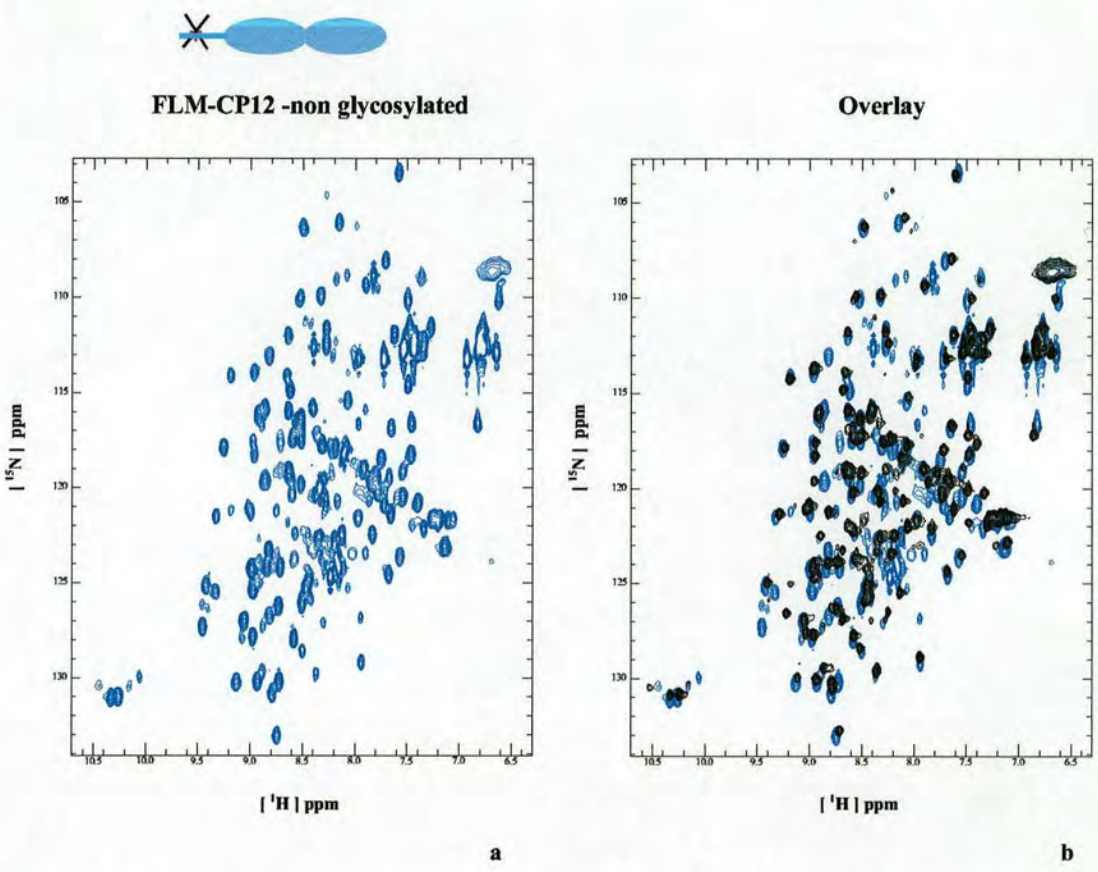



**Figure 4.6: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) FLM-CP12 -non glycosylated**

Sample was 0.7 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$.

Spectrum was acquired at 600 MHz, and 37°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2212.6 Hz in the $^{15}$N dimension. Spectrum was recorded with 10 scans, 1024 complex points in the $^{1}$H dimension, and 128 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 128 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) TR-CP12 -non glycosylated (black) and FLM-CP12 – glycosylated (blue) overlaid**

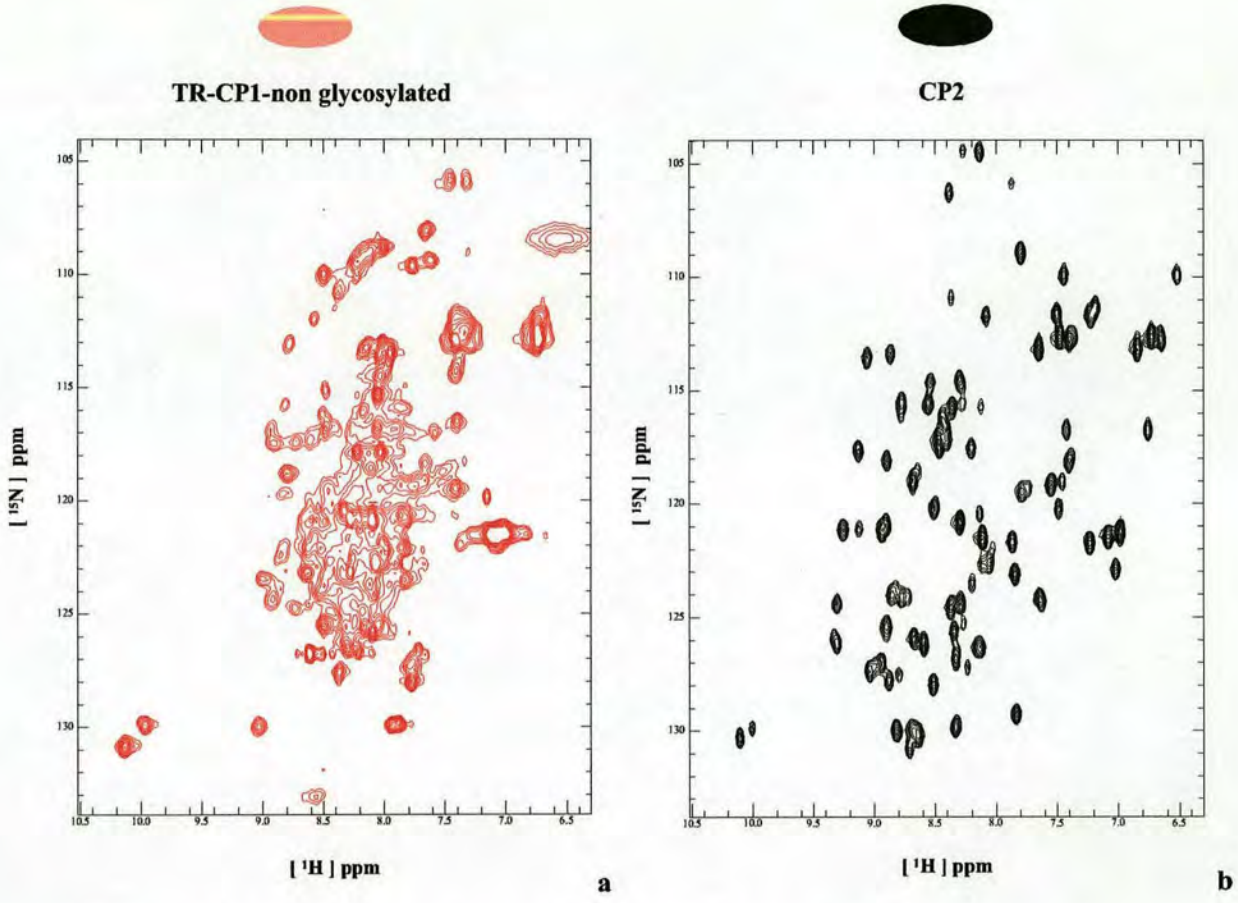



**Figure 4.7: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) TR-CP1 -non glycosylated**

Sample was 0.7 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 25°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2000 Hz in the $^{15}$N dimension. Spectrum was recorded with 8 scans, 1024 complex points in the $^{1}$H dimension, and 192 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 256 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) CP2**

Sample was 0.7 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 25°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2000 Hz in the $^{15}$N dimension. Spectrum was recorded with 8 scans, 1024 complex points in the $^{1}$H dimension, and 128 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 128 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

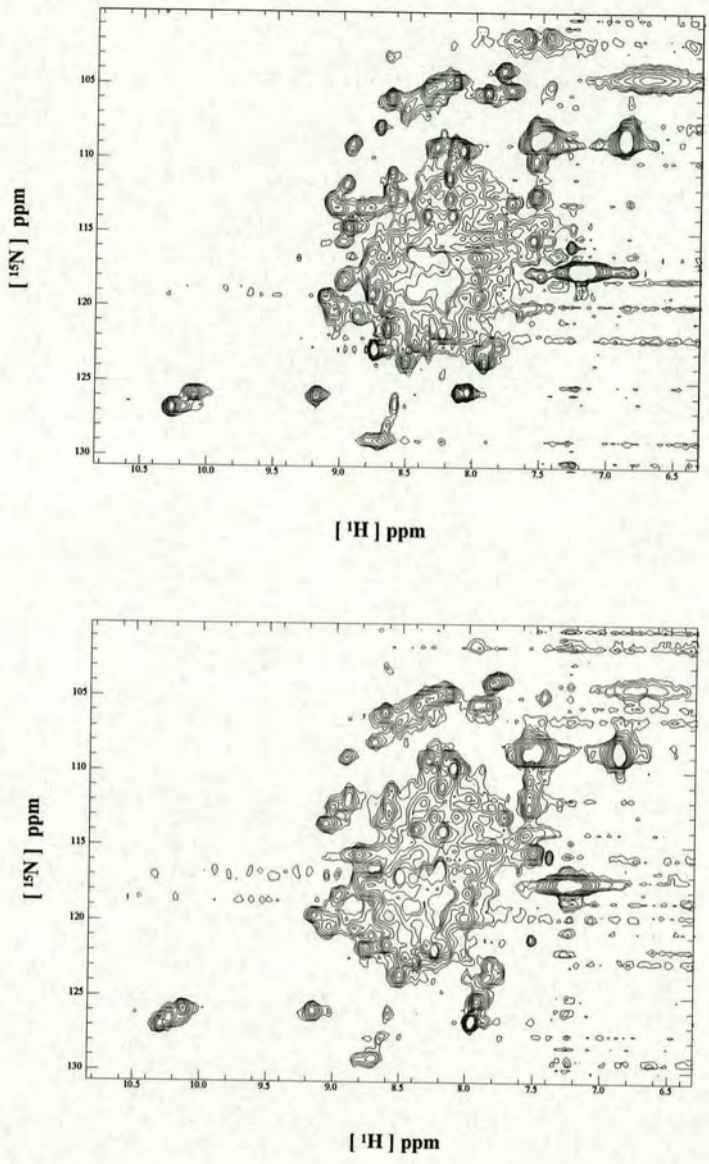


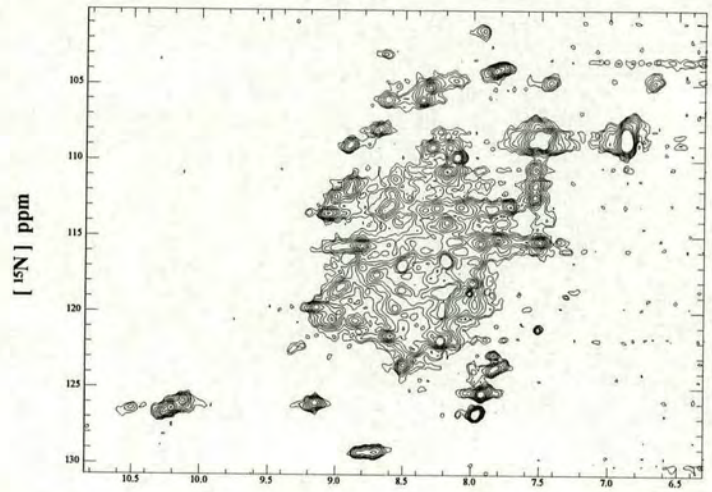

**Figure 4.8: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) TR-CP1 -non glycosylated at pH 4**

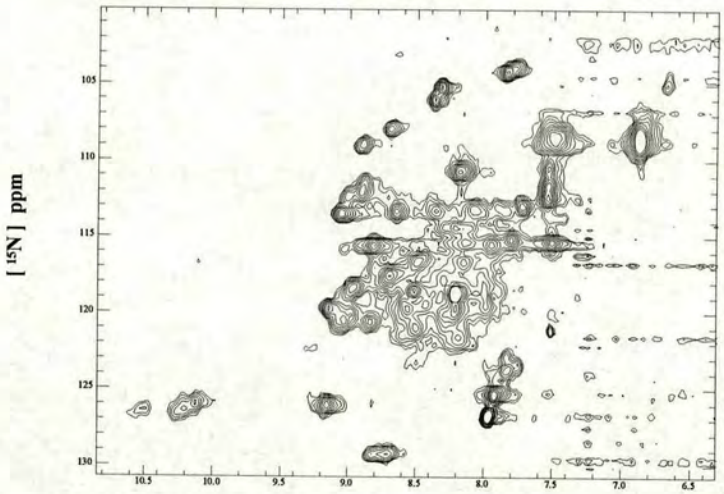
Sample was 0.8 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 25°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2216.6 Hz in the $^{15}$N dimension. Spectrum was recorded with 8 scans, 1024 complex points in the $^{1}$H dimension, and 96 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 128 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) TR-CP1 -non glycosylated at pH 6**

Same experimental conditions as in (a), pH was adjusted with NaOH.




c




d

**Figure 4.8 continued: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**c) TR-CP1 -non glycosylated at pH 7.5**

Same experimental conditions as in (a), pH was adjusted with few ml of 0.2 M NaOH.

**d) TR-CP1 -non glycosylated at pH 8.5**

Same experimental conditions as in (a), pH was adjusted with few ml of 0.2 M NaOH.

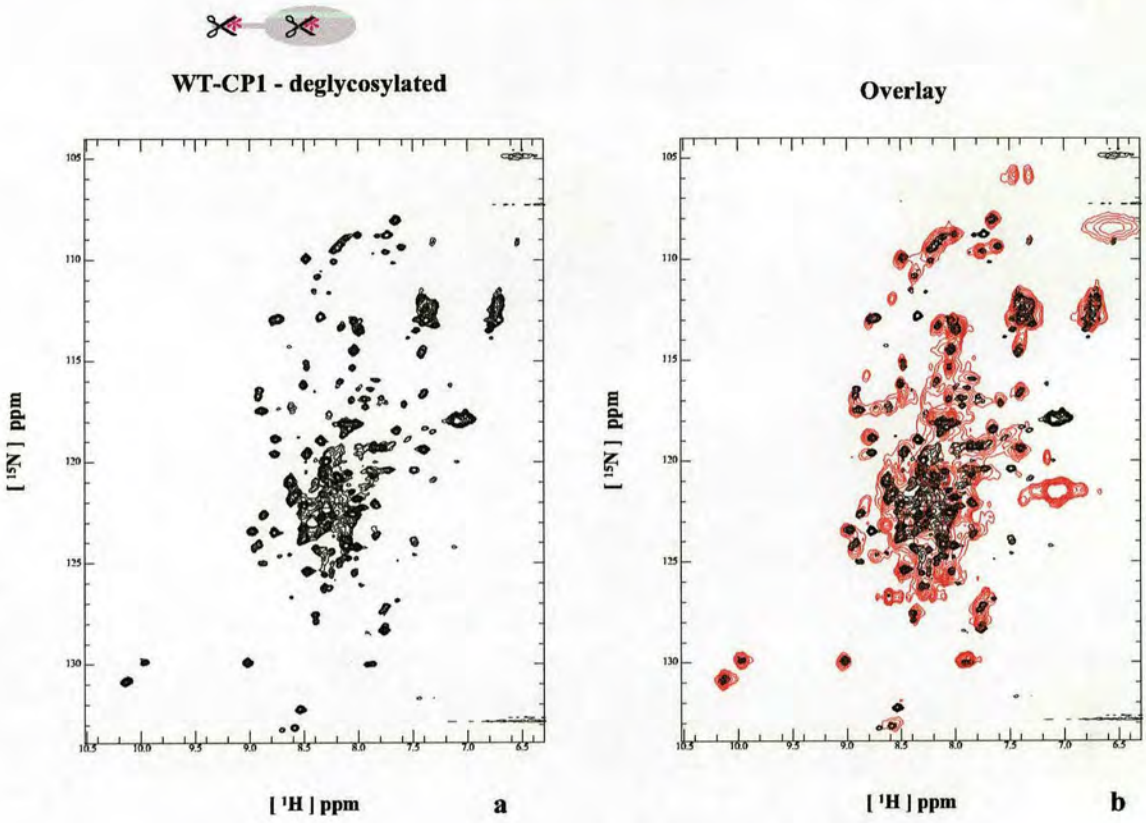



**Figure 4.9: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) WT-CP1-deglycosylated**

Sample was 0.6 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 25°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2000 Hz in the $^{15}$N dimension. Spectrum was recorded with 64 scans, 1024 complex points in the $^{1}$H dimension, and 160 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 256 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) WT-CP1-deglycosylated (black) and TR-CP1 -non glycosylated (red) overlaid**

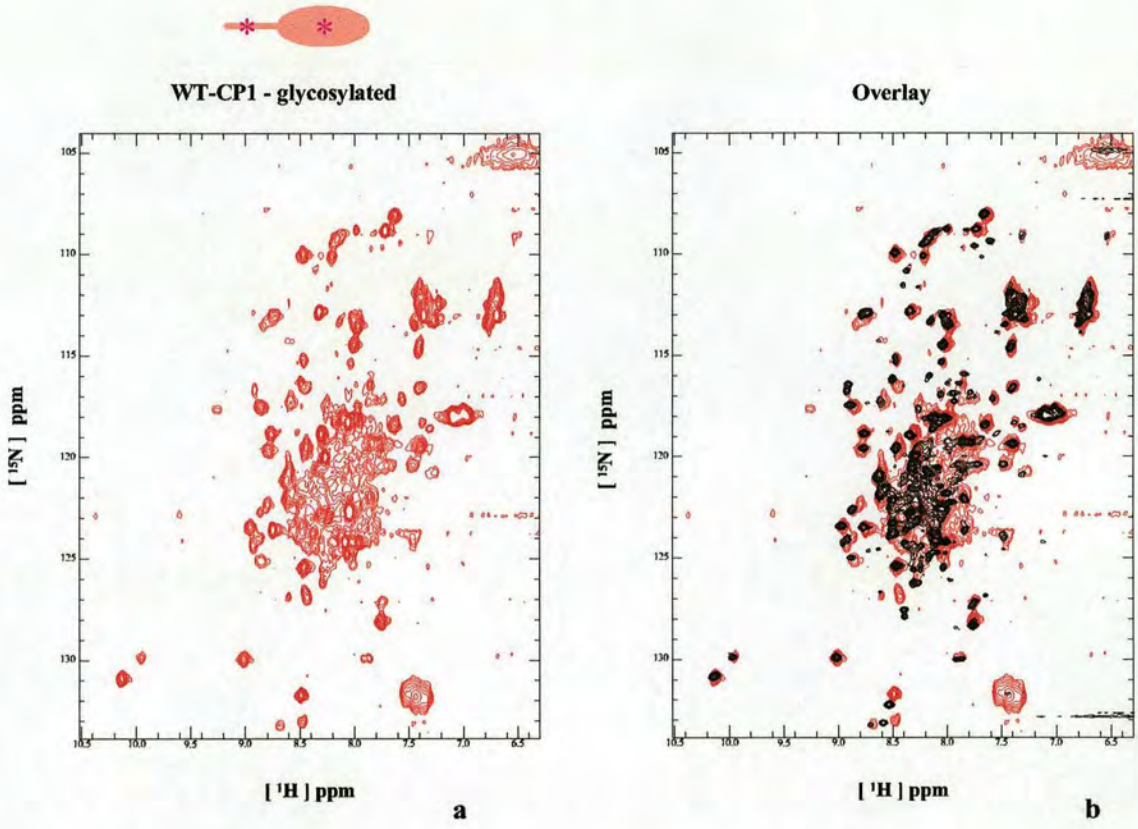



**Figure 4.10: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) WT-CP1 -glycosylated**

Sample was 0.7 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 25°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2000 Hz in the $^{15}$N dimension. Spectrum was recorded with 16 scans, 1024 complex points in the $^{1}$H dimension, and 96 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 128 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) WT-CP1 -deglycosylated (black) and WT-CP1 – glycosylated (red) overlaid**

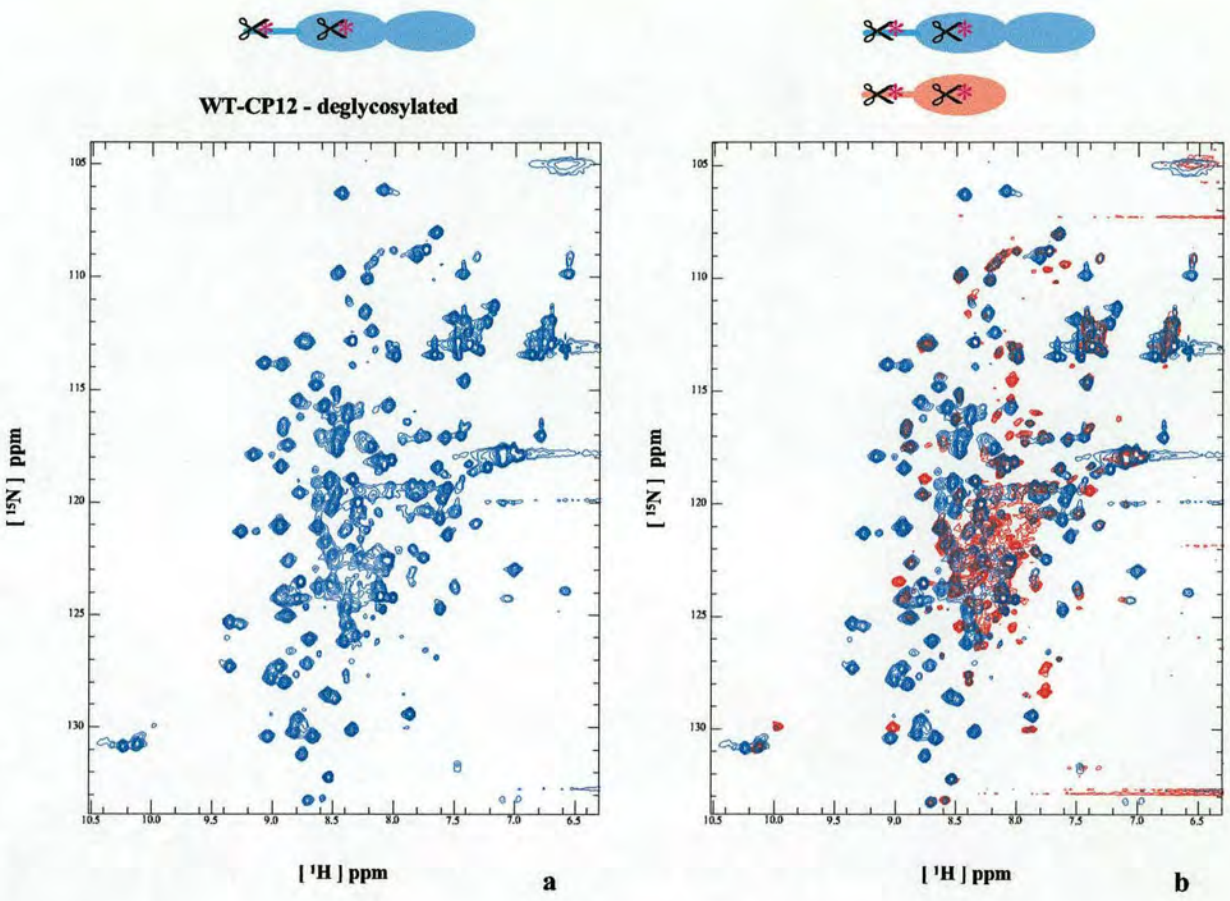



**Figure 4.11: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) WT-CP12 -deglycosylated**

Sample was 0.7 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 25°C. Sweep width is 5000 Hz in the $^{1}$H dimension and 2000 Hz in the $^{15}$N dimension. Spectrum was recorded with 16 scans, 1024 complex points in the $^{1}$H dimension, and 96 complex points in the $^{15}$N dimension. After processing, data matrix was 1024 points in $^{1}$H, and 128 points in $^{15}$N. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) WT-CP12 -deglycosylated (blue) and WT-CP1 – deglycosylated (red) overlaid**

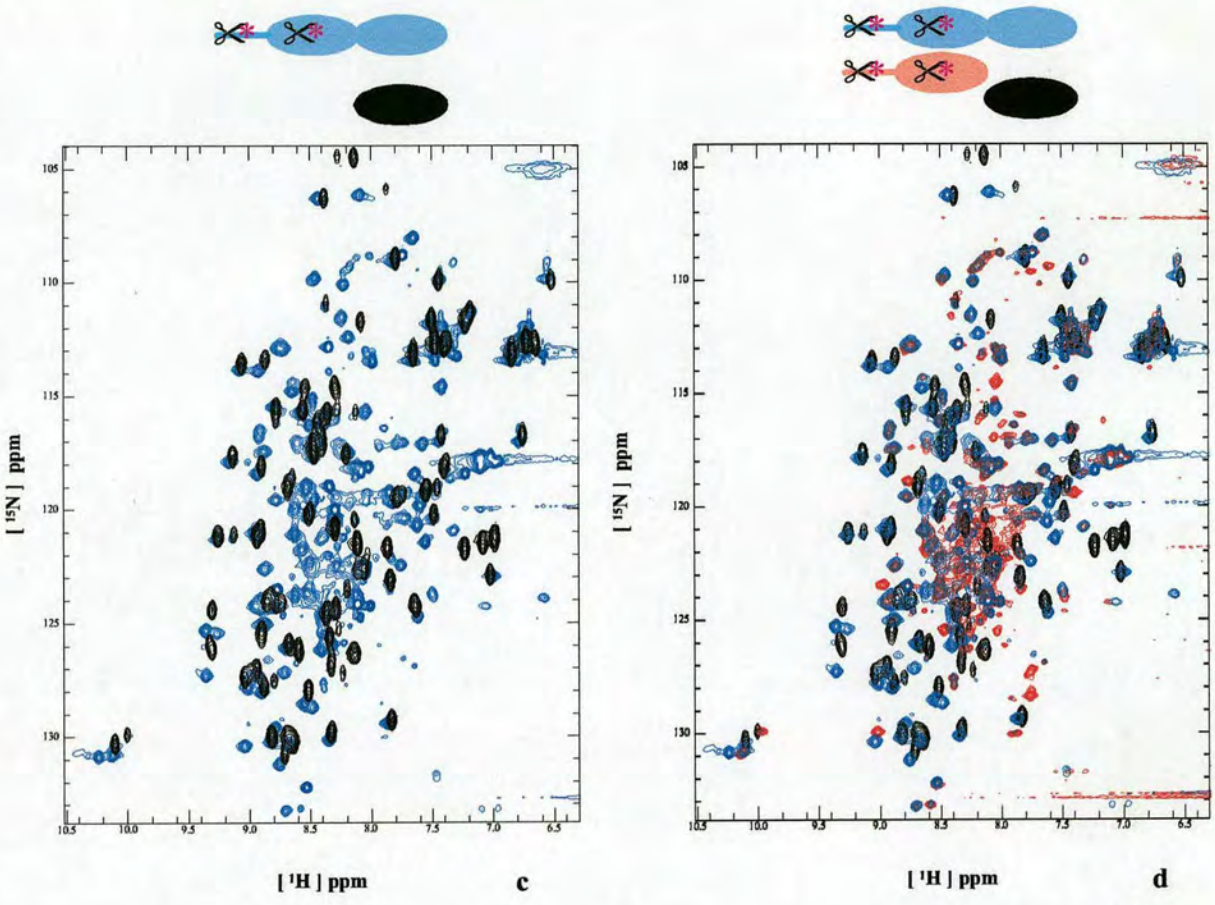



**Figure 4.12: $^{1}$H-$^{15}$N HSQC-spectra of GABA$_B$ R1a CP-fragments**

**a) WT-CP12 -deglycosylated (blue) and CP2 (black) overlaid**

**b) WT-CP12 -deglycosylated (blue), WT-CP1 – deglycosylated (red) and CP2 (black) overlaid**

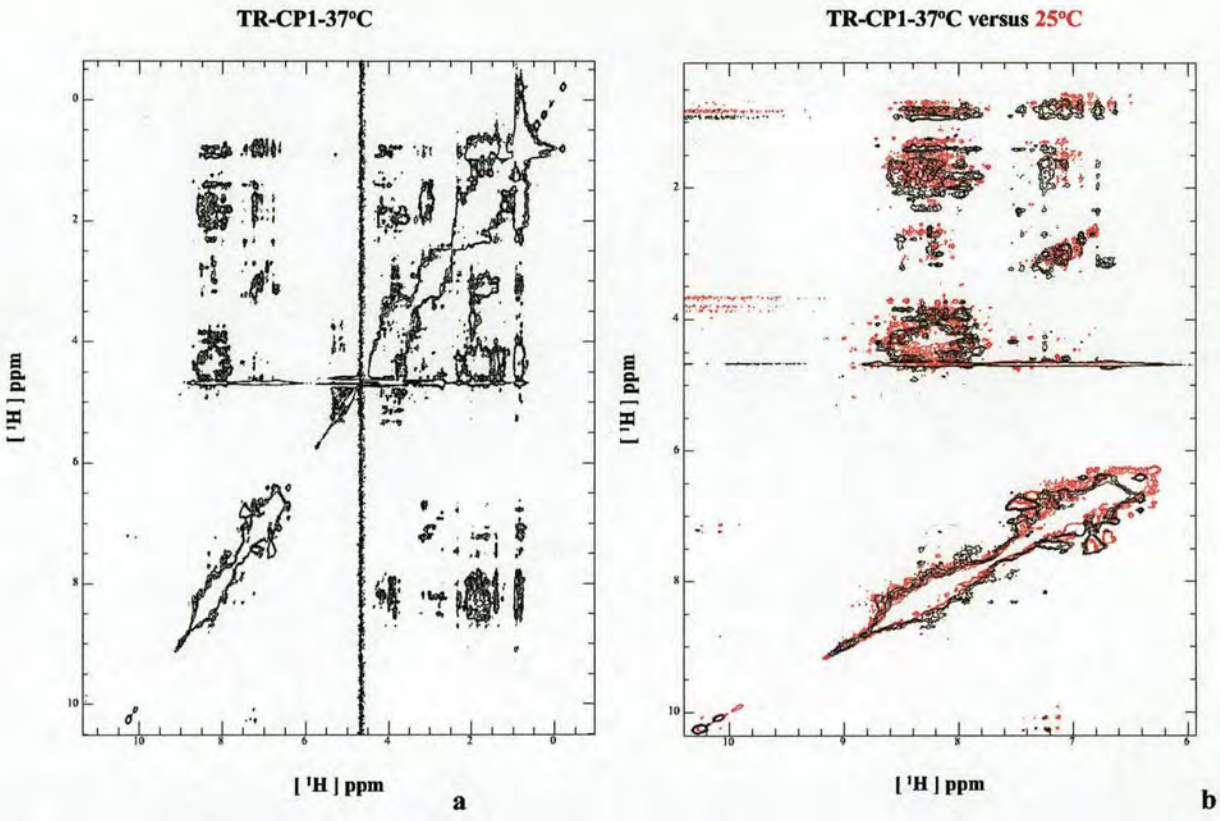



**Figure 4.13: $^1$H-2D NOESY-spectra of $GABA_B$ R1a CP-fragments**

**a) TR-CP1-non glycosylated**

Sample was 0.7 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 37°C. Sweep width is 8000 Hz in the acquisition dimension and 7000.35 Hz in the indirectly detected dimension. Mixing time was 0.1 s. Spectrum was recorded with 32 scans, 2048 complex points and 1024 complex points in the indirectly detected dimension. After processing, data matrix was 2048 points in the acquisition dimension and 1024 points in the indirectly detected dimension. A sine function squared (with an angle of 90°) was applied in both dimensions.

**b) TR-CP1-non glycosylated at 37°C (black) and TR-CP1-non glycosylated at 25°C (red)**

Spectrum recorded at 25°C was acquired at 600 MHz. Sample conditions, spectrum recording and processing was as in (a).

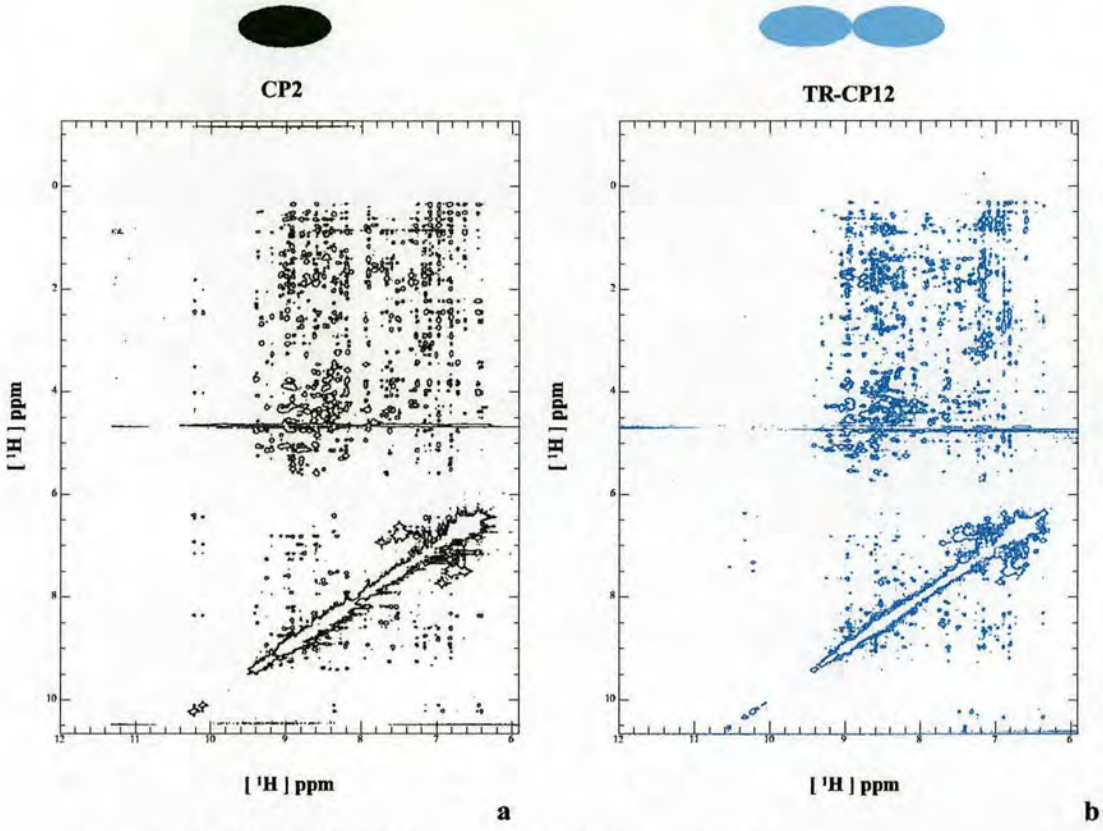



**Figure 4.14: $^{1}$H-2D NOESY-spectra of $GABA_B$ R1a CP-fragments**

**a) CP2**

Sample was 0.7 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 37°C. Sweep width is 8000 Hz in the acquisition dimension and 7000.35 Hz in the indirectly detected dimension. Mixing time was 0.1 s. Spectrum was recorded with 92 scans, 2048 complex points and 1024 complex points in the indirectly detected dimension. After processing, data matrix was 2048 points in the acquisition dimension, and 1024 points the indirectly detected dimension. A sine function squared was applied in both dimensions.

**b) TR-CP12-non glycosylated**

Sample was 0.6 mM in 20 mM deuterated sodium acetate, pH 4, and 10% $D_2O$. Spectrum was acquired at 600 MHz, and 37°C. Sweep width is 10000Hz in the acquisition dimension and 7199.42 Hz in the indirectly detected dimension. Mixing time was 0.1 s. Spectrum was recorded with 80 scans, 2048 complex points and 1024 complex points in the indirectly detected dimension..After processing, data matrix was 2048 points in the acquisition dimension, and 1024 points the indirectly detected dimension. A sine function squared was applied in both dimensions.

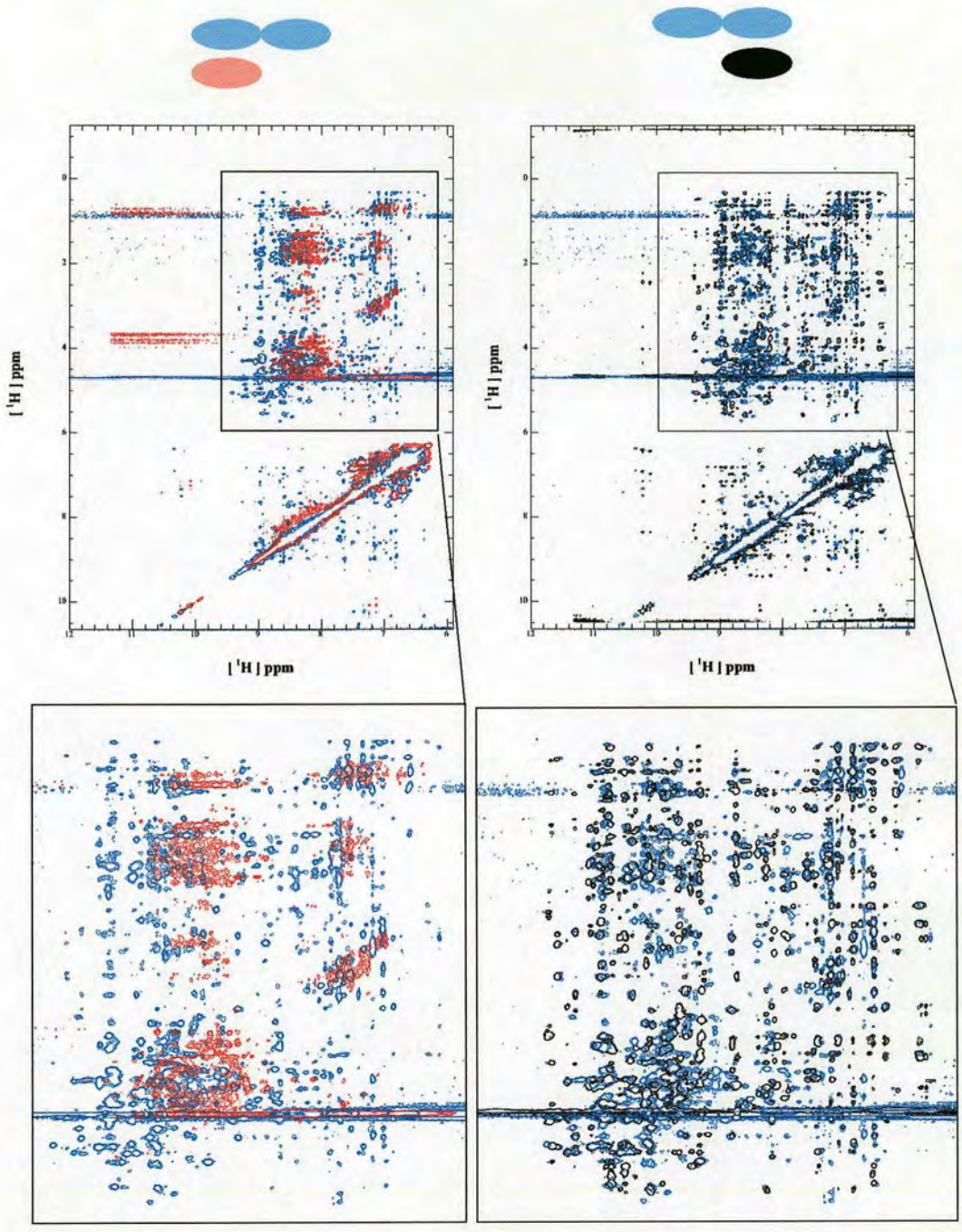



**Figure 4.15: $^{1}$H-2D NOESY-spectra of GABA$_B$ R1a CP-fragments overlays**

**Left: TR-CP1 at 25°C (red) & TR-CP12 at 37°C (blue)**

**Right: CP2 at 37°C (black) & TR-CP12 at 37°C (blue)**

## 4.2 Circular dichroism studies

Far-UV (180-250 nm) circular dichroism (CD) spectra of proteins are dominated by peptide bond dichroism, and therefore are used to estimate their secondary structure content [176-178]. The spectrum of a protein is basically the sum of the spectra of its conformational elements. α-helices display large CD bands with negative ellipticity at 208-222 nm, and positive ellipticity at 193 nm. β-sheets exhibit a broad negative band near 218 nm and a large positive band near 195 nm. Finally, random coils have a weak positive broad band at 217 nm and a large negative band near 200 nm.

Far-UV CD spectra of different constructs of the $GABA_B$ R1a CP modules were obtained at various temperatures. These CD data were collected during the course of testing different pH and buffer conditions for NMR spectroscopy; hence, the CD data of each construct was recorded with different sample conditions, and a direct comparison of these results is difficult. Nevertheless these CD data are indicative of the secondary structure content, and complement the NMR data shown in section 4.1.

Far-UV CD spectra of CP modules are characterised by a highly unusual large positive ellipticity in the 220-260 nm region (Figure 4.16 a-d) [34, 179], which could arise from the conserved tryptophan residue being in a β-sheet environment [179]. All $GABA_B$ R1a complement module pairs or single module constructs display this unusual characteristic, which is therefore consistent with a CP module fold for both modules in the $GABA_B$ R1a. Changes in ellipticity of a protein as a function of temperature are a convenient way of probing the thermodynamics of folding [176]. Thermal denaturation as monitored by CD was therefore used to probe folding and stability of the various constructs. Similar experiments on well characterised CP module pairs or single modules have shown that they unfold in a cooperative manner [179], i.e. with a concerted loss of the interactions stabilising the native structure, which can be described by a "two state" model. Figure 4.17 a shows the thermal unfolding of the different constructs of the $GABA_B$ R1a CP module pair and single modules along with a well characterised single CP module ($16^{th}$ module of CR1), and a CP-module pair (from the viral protein VCP: VCP-$2^{nd}$ and $3^{rd}$ module (VCP23)). CP2, and VCP23 show clear sigmoidal transitions, consistent with cooperative unfolding, and the CR1-$16^{th}$ transition is also sigmoidal. On the other hand WT-CP1 and FL-CP12 have less well defined transitions which seem be consistent with little or no cooperativity in the unfolding process. The latter has a profile close to the sum of its component $GABA_B$ R1a single module-profiles.

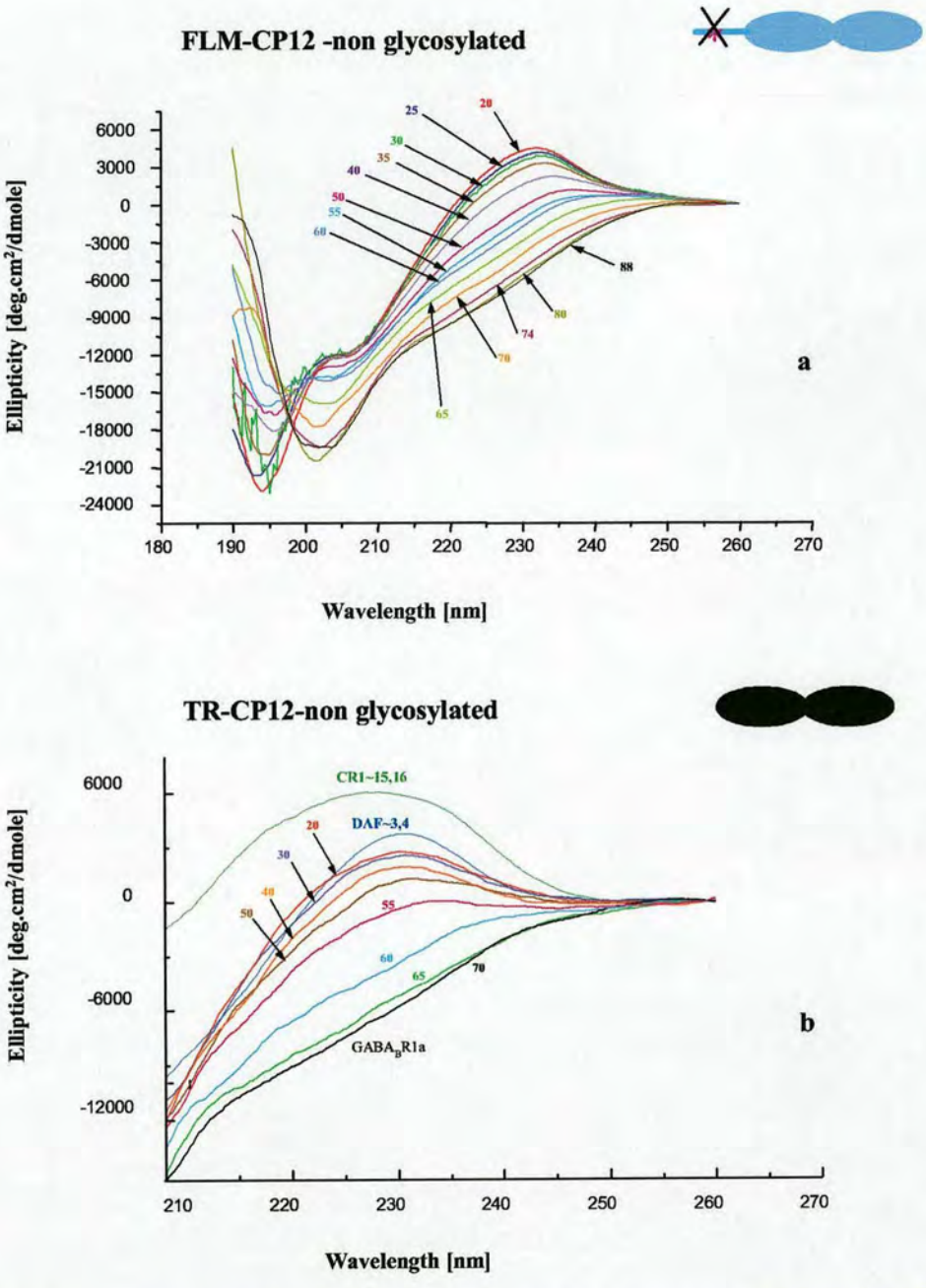



**Figure 4.16: Thermally induced denaturation of the $GABA_B$ R1a CP fragments as monitored by far-UV CD.** CD spectra were obtained on a Jasco-600 spectropolarimeter (Japan Spectroscopic Co, Tokyo, Japan), using a cylindrical quartz cell with a path length of 0.5 cm for (a), and (b) measurements. Temperature is indicated along individual spectrum, and is given in °C.

**a) FLM-CP12, 50 μM in 20 mM sodium phosphate pH 6**

**b) TR-CP12, 30 μM in 25 mM sodium phosphate pH 8.** ***Taken from reference [34]***

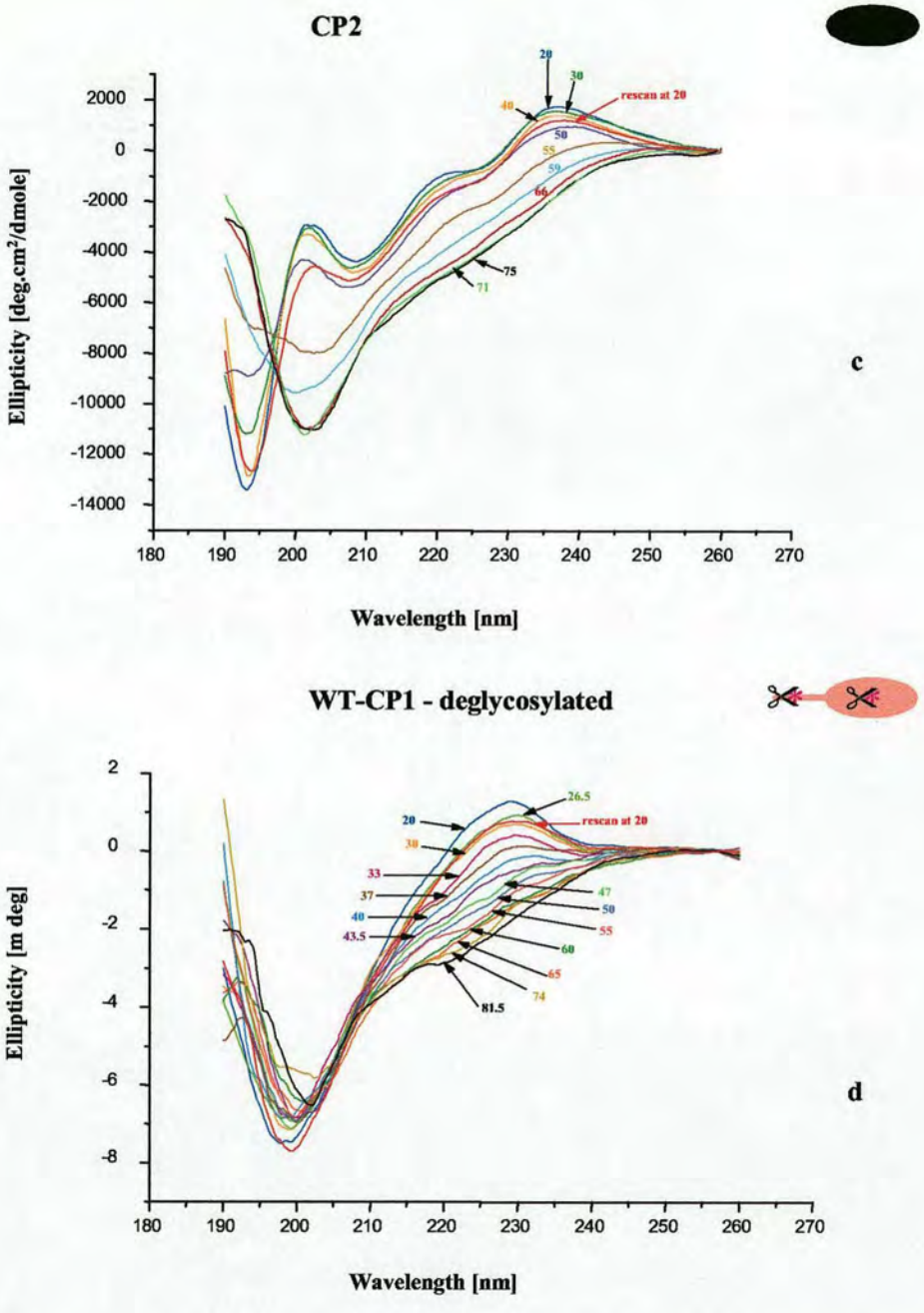



**Figure 4.16 continued: Thermally induced denaturation of the $GABA_B$ R1a CP fragments as monitored by far-UV CD.** CD spectra were obtained on a Jasco-600 spectropolarimeter (Japan Spectroscopic Co, Tokyo, Japan), using a cylindrical quartz cell with a path length of 0.5 cm for (c) measurement, and of 0.02 cm for the (d) measurement. Temperature is indicated along individual spectrum, and is given in °C.

**c) CP2, 80 μM in 20 mM sodium phosphate pH 6**

**d) WT-CP1-deglycosylated, 50 μM in 20 mM deuterated sodium acetate pH 4 (NMR buffer)-** Note: these measurements are uncorrected for cell path length and protein concentration.

Aggregation was suspected from $^{15}N$ $T_2$ experiments, and could explain the poor cooperativity in unfolding observed for WT-CP1 and FLM-CP12. However it worth noting that protein concentration in these experiments is relatively low (0.05 mM).

The near-UV (260-320 nm) CD of proteins arises from the environment of each aromatic amino acid side-chain. In the case of structured proteins in their native states, the side-chains of such amino acids are placed in an asymmetric environment that is characteristic of the tertiary structure of the protein. The tryptophan contribution gives rise to a peak close to 290 nm with a fine structure between 290 and 305 nm. Tyrosine's contribution is a peak between 275 and 282 nm, and phenylalanine gives sharp peaks between 255 and 270 nm. Disulphide bond contributions are normally seen at around 260 nm, although their CD bands are usually much broader than those of aromatic amino acids and in general it is difficult to evaluate their contribution to the overall spectrum (Dr. Sharon Kelly - Glasgow University - personal communication). Disulphide bond can also contribute to the far-UV CD at 225 nm.

For the WT-CP1, the near-UV spectrum shows a "u-shaped" band with negative ellipticity between 260-310 nm, and a minimum around 280 nm (Figure 4.17 b). A similar profile was observed in the case of CR1~15-17 fragment (Figure 4.17 c). Despite the fact that in the latter, three modules contributed to the spectra; it is still possible to speculate that such a profile is characteristic of CP modules in the near-UV CD. A fine structured "shoulder" is also observed at 295 nm for both WT-CP1 and CR1~15-17, which might represent the contribution from the conserved tryptophan. Denaturation monitored by near-UV CD will be needed to further characterise the extent of these tertiary interactions. It also will be of interest to repeat near-UV CD measurement on reduced and carboxymethylated CP1, to further characterise disulphide bonds contributions to the spectra.

Far-UV CD data for WT-CP1 are consistent with an unstable protein, which exhibits little if any cooperativity in unfolding, a feature generally found in molten globules. Data from near-UV CD are consistent with the presence of some native or native-like tertiary structure. Interestingly, reduction of the disulphide bonds was shown to abolish the characteristic far-UV positive ellipticity of the TR-CP12 construct [34]. This result implies an important disulphide bond contribution to the CD spectra of the $GABA_B$ R1a ~CP modules and possibly of CP modules in general.

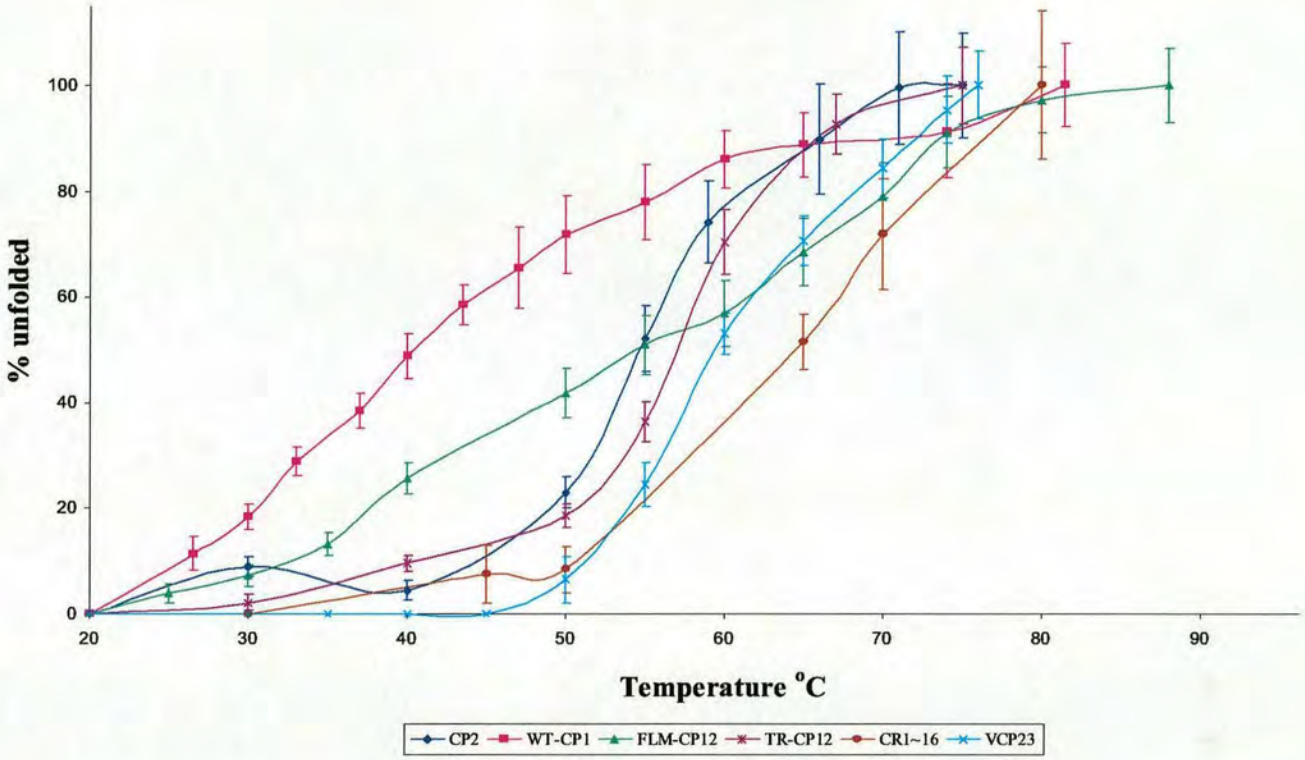



**Figure 4.17: CD analysis of the $GABA_B$ R1a CP fragments.** CD spectra were obtained on a Jasco-600 spectropolarimeter (Japan Spectroscopic Co, Tokyo, Japan).

**a) Thermally induced denaturation of the $GABA_B$ R1a fragments and others CP fragments from human complement or Vaccina virus complement control protein monitored by far-UV CD.** Plots of the percentage total change versus temperature were derived from measurements shown in figure 4.16 by taking the maximum ellipticity. CR1~16 [179] and VCP23 values were extracted from reference [180]. The observed change from the native protein (0 M GdnHCl) at each concentration was expressed as a percentage of the total change induced by 6 M GdnHCl, and plotted against the concentration of GdnHCl. The maximum ellipticity and readings at ±1.0 and ±2.0 nm of the maximum were averaged to smooth experimental errors in the curves.

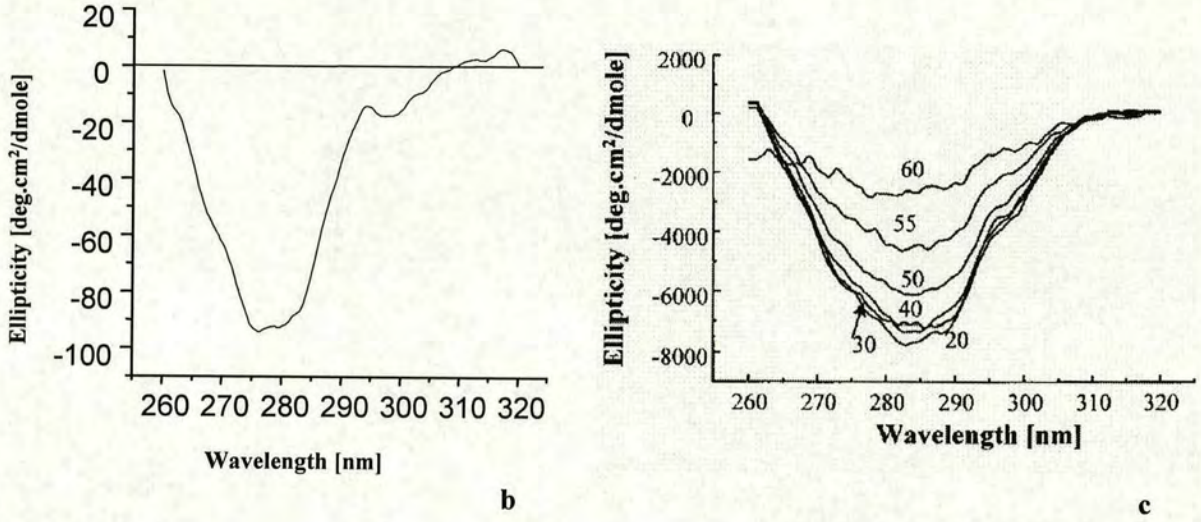



**Figure 4.17 continued: CD analysis of the $GABA_B$ R1a CP fragments.** CD spectra were obtained on a Jasco-600 spectropolarimeter (Japan Spectroscopic Co, Tokyo, Japan).

**b) Near-UV CD spectrum of WT-CP1 - deglycosylated – 51 μM in 20 mM deuterated sodium acetate pH 4 (NMR buffer) - cell path-length was 0.02 cm.**

**c) Near-UV CD spectra for CR1~15-17 over a range of temperature, as indicated by the numbers.*Taken from reference [179]***

## 4.3 Differential scanning calorimetry studies

The temperature-induced denaturation of a protein enables assessment of stability. Differential scanning calorimetry (DSC) can be used to analyse the thermodynamics of protein unfolding. With this technique one measures, as a function of temperature, the excess heat capacity ($C_p$) of a protein-containing solution, relative to the same volume of a buffer solution [181]. After concentration and baseline correction, the DSC data are fitted to a model of unfolding to yield the apparent thermodynamic parameters for the transition. For protein unfolding transitions, the model can be a "two-state" model or a multi-state process. The transition enthalphy ($\Delta H$) is obtained by integration of the experimental curve of $C_p$ versus temperature. This model-free measure of $\Delta H$ is then compared with the van't Hoff transition enthalphy ($\Delta H_V$), which is calculated from the shape of the modeled DSC curve and is a measure of the cooperativity of unfolding: the sharper the transition, the greater the $\Delta H_V$, and the greater the cooperativity [181]. Such a comparison gives information about the thermodynamics of the order-disorder transition observed. If the model of unfolding is correct then $\Delta H = \Delta H_V$; if not, discrepancies between the two enthalpies are of interest [182]. For example, $\Delta H < \Delta H_V$ can arise from dimerisation, aggregation, or from not all the protein being correctly folded in the first place [183]. The melting temperature ($T_m$) refers to the temperature at which 50% of the protein molecules are unfolded [183].

Figure 4.18 shows the profiles obtained from a DSC study of the $GABA_B$ R1a~CP fragments. The DSC profile of CP2 (Figure 4.18 a) was fitted with a "non two-state" unfolding model giving a good agreement between $\Delta H$ and $\Delta H_V$. The profile is a sharp with a relatively high melting temperature (70.8°C) as expected for a protein that has a stable and well-defined tertiary structure (cooperative unfolding). In comparison the DSC profile obtained for WT-CP1 (Figure 4.18 b) shows no transition, indicating that endothermic unfolding did not occur, i.e. the fragment absorbed no excess energy. A similar result was found in our laboratory with the *Herpes simplex* virus triplex protein (VP23) [184]. VP23 was found to exist as a molten globule, and absence of endothermic unfolding transition was also observed, consistent with the lack of rigid tertiary structure generally observed in the molten globule state. The absence of endothermic unfolding as monitored by DSC for WT-CP1 does not agree with its near-UV CD data that is suggestive of the presence of tertiary structure. Aggregation could provide an explaination for the lack of endothermic transition for WT-CP1, and this experiment will need to be repeated at a more dilute protein concentration.

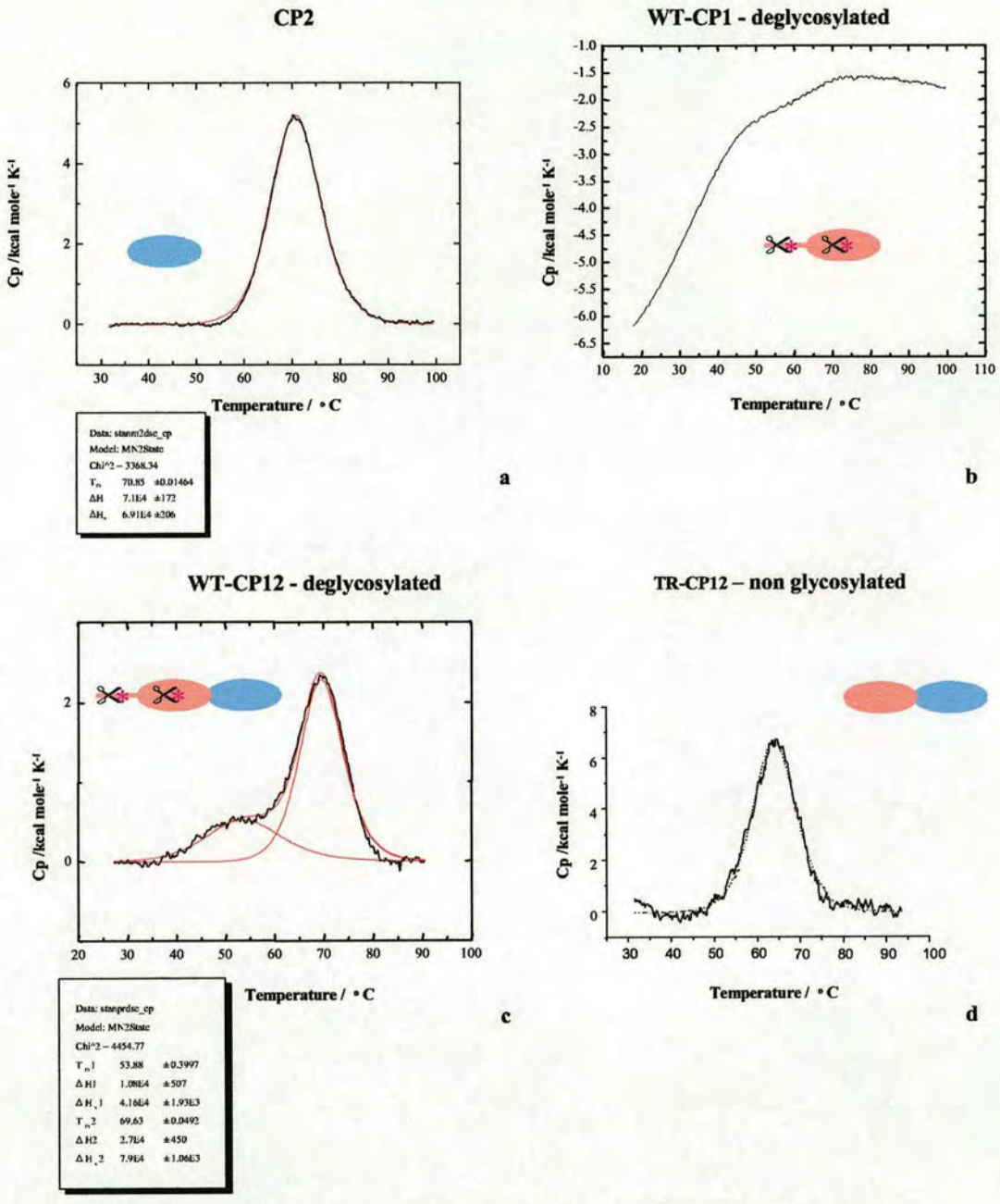



**Figure 4.18: Differential scanning calorimetry studies on $GABA_B$ R1a CP fragments.**

Studies were performed on a MC-2 differential scanning calorimeter (Microcal, Northampton, MA-USA). The cell volume was 1.5 ml, heating rate was 1°C $min^{-1}$, and excess of pressure was kept equal to 80 bar.

**a) CP2, 40 μM in 20 mM sodium acetate pH 4.**

**b) WT-CP1 - deglycosylated, 76 μM in 20 mM sodium acetate pH 4.**

**c) WT-CP12 - deglycosylated, 40 μM in 20 mM sodium acetate pH 4.**

**d) TR-CP12 – non glycosylated- 100 μM in 25 mM sodium phosphate pH 8.** ***Taken from reference*** **[34].**

Interestingly, the DSC profile for WT-CP12 (Figure 4.18 c) shows two transitions. These data were also fitted with a "non two-state" unfolding transition yielding discrepancy between the two enthalpy: $\Delta H < \Delta H_V$ in both observed transitions. Such discrepancies can arise from several factors; one of them is incorrect baseline correction. Baseline correction is achieved subtracting a blank, recorded with buffer solution in both measurement and reference cells, to the measurement. The WT-CP12 sample used in this study was buffer-exchanged using an ultra-filtration concentrator. Ultra-filtration does not lead to a full equilibrium between sample and buffer (Dr Alan Cooper - Glasgow University - personal communication), and it would have been better to use dialysis (note: dialysis was employed for both WT-CP1 and CP2). Nevertheless the major transition has a $T_m$ close to that of the CP2 fragment (69.6°C) and could account for the melting of the second module in the module-pair fragment. The first transition, with a $T_m$ of 53°C, (if real) could be accounted by an aggregation effect or from the melting of the seven-residue extension. In a previous study [34], Dr Marina Kirkitadze found only one transition for the thermal unfolding of the TR-CP12 fragment with a $T_m$ of 64°C (see Figure 4.18 d). It is therefore tempting to speculate that the second transition arises form the seven-residue extension. Alternatively, this lower transition could be accounted for by CP1 if it was more structured in the context of the module pair. New measurements on dialysed samples will be the focus of future work.

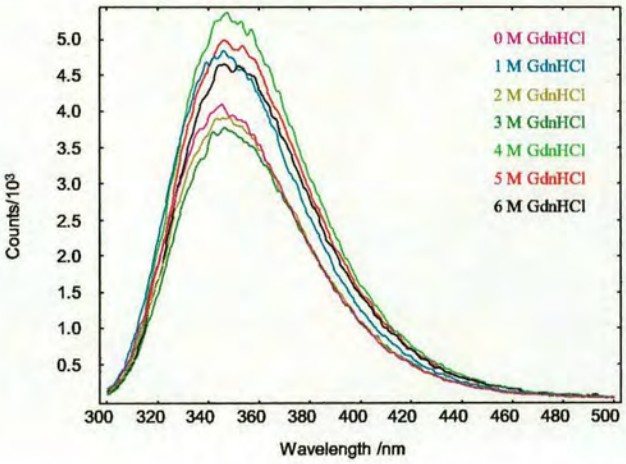
## 4.4 Fluorescence spectroscopy studies

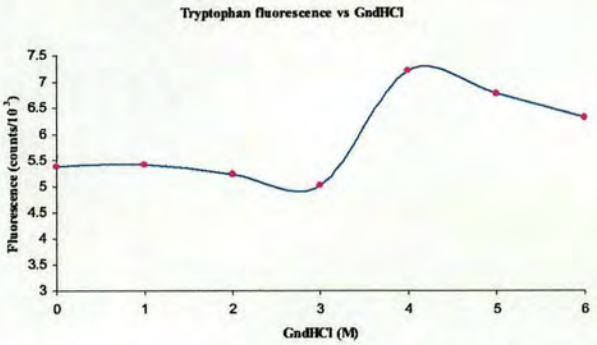
The $GABA_B$ R1a~ CP1 has a conserved tryptophan, as well as an extra tryptophan residue in the predicted first loop. The CP module consensus structure has been well characterised and the conserved tryptophan side-chain is known to be mainly buried within the hydrophobic core of the module, adjacent to the first disulphide bond and close to the N-terminus. Fluorescence spectroscopy can be used to assess the extent to which tryptophan side-chains are buried within a protein interior. Excitation at 295 nm results in a tryptophan fluorescence emission spectrum with a maximum at 340 nm [185]. Upon denaturation, this maximum is in general shifted towards 350-355 nm; however, it is difficult to predict how the intensity will change, and both increased and decreased intensity have been observed upon denaturation [185]. An increase in fluorescence emission intensity at 340 nm upon unfolding has been observed for some CP modules of complement receptor 1 (CR1) [179, 186]. CR1~$16^{th}$ module showed a seven-fold increase in fluorescence emission intensity at 340 nm in 7 M GdnHCl, as well as a shift in the maximum to 351 nm.

Protein fluorescence measurements on WT-CP1 showed that as the GdnHCl concentration was increased, a buried tryptophan side-chain became exposed to the solvent; presumably this is the side-chain of the conserved tryptophan residue (Figure 4.19 a). Conclusions from fluorescence unfolding experiments of proteins containing multiple tryptophan residues are difficult to draw, but the main transition appears to be real and is likely to correspond to the exposure of the buried side-chain of the conserved tryptophan residue. Interestingly, CR1~15-16, CR1~16-17, and CR1~16 fluorescence unfolding measurements all showed a transition at 3.3-3.6 M GdnHCl [179, 186]; and are consistent with the main transition observed for the $GABA_B$ R1a~ WT-CP1.

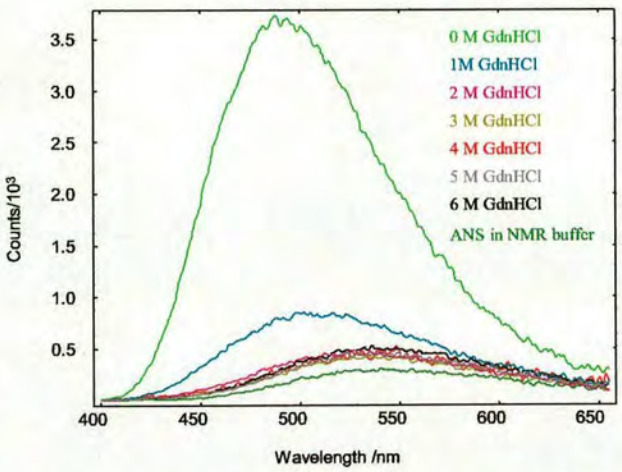
However, examples of molten globule where a tryptophan side-chain is hidden from the solvent have been observed. As mentioned earlier, *Herpes simplex* virus triplex protein (VP23) was found to exist as a molten globule [184] with an unfolding transition detected by fluorescence taking place around 3 M GdnHCl. The VP23 transition was found to be poorly cooperative, which was attributed to the lack of defined tertiary structure. Native-like or residual tertiary structure could therefore account for the transition observed in WT-CP1.

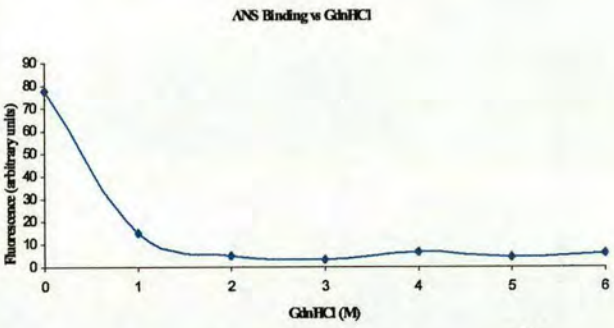
A sensitive test to probe partial folding is binding of 8-anilino-1-naphtalenesulfonate (ANS) to a protein. Binding of ANS induces an increase in fluorescence at 470 nm following excitation at 370 nm [184].




**a**




**b**

**Figure 4.19: Fluorescence studies on WT-CP1**

Fluorescence measurements were recorded on an Edinburgh analytical instrument-F900 spectrofluorimeter in a 100 µl cuvette of 0.3 cm path length at 26.2°C. The protein concentration in all experiments was 71 µM (0.67 mg/ml) in NMR buffer (20 mM deuterated sodium acetate, pH 4).

**a) Protein fluorescence: excitation wavelength was 295 nm and the emission spectra were recorded between 300 and 500 nm.**

**b) ANS fluorescence: ANS was purchased from Molecular Probes, and was added to protein samples to a final concentration of 10 µM.** Plots were obtained from data integration corrected from free ANS in NMR buffer value.

ANS binds to the exposed hydrophic regions in partially folded proteins, although ANS has also been shown to bind to solvent-accessible non-polar groups in native proteins; binding to a partially folded state is in general much stronger compared to the native or fully denatured state [184].

ANS was found to bind tightly to purified, undenatured WT-CP1 (Figure 4.19 b). Measurements of the unfolding dependence of ANS fluorescence intensity at 470 nm showed that as the GdnHCl concentration was increased, fluorescence intensity decreases rapidly. They also showed that the fluorescence of ANS in the absence of WT-CP1 or in the presence of its denatured form in 6 M GdnHCl were insignificant. ANS fluorescence intensity was rapidly quenched upon addition of GdnHCl; most of the intensity was quenched by 1 M GdnHCl and it was completely lost at 2 M GdnHCl. In a recent study, Bailey *et al* [187] showed that clusterin, a ubiquitous protein with several binding partners, exists as a molten globule. In this study, clusterin and other proteins along with their pH-induced molten globules were compared in their ability to bind ANS. Two proteins (apomyoglobin and α-lactalbumin) were found to bind ANS in either their native or molten globule states (with different intensities). However, different unfolding profiles as monitored by ANS binding were found for the native and molten globule states. Native state unfolding profiles for both proteins were sigmoidal and consistent with cooperative unfolding. On the other hand, unfolding profiles for their molten globule states were non-sigmoidal and therefore consistent with non-cooperative unfolding as previously observed for other molten globules. Furthermore native states required higher denaturant concentration to induce ANS fluorescence loss compare to molten globule states; complete loss was observed at 2 M urea for molten globule states while 3-6 M was needed for native proteins [187]. The profile obtained for the $GABA_B$ R1a~ WT-CP1 is similar to clusterin and others molten globules described in Bailey's study; more data points will be needed between 0-1 M GdnHCl to further characterise such a profile.

## 4.5 Light scattering and analytical centrifugation

The poor quality of CP1 NMR data could be accounted for by the protein being an aggregate, despite the lack of improvement upon dilution. Early in this study, samples of the $GABA_B$ R1a~ TR-CP12 at NMR concentration (1 mM) were submitted to different collaborators for molecular mass determination. Dr. Kormelia Jumel performed analytical centrifugation studies at Nottingham University in Dr. Stephen Harding's laboratory, and Dr. Jacqueline Dornan at the Institute of Cell and Molecular Biology in Edinburgh University, performed light scattering studies.

Analytical ultracentrifugation is a powerful tool with which to measure a protein's molecular weight, shape and association properties. Two types of experiments can be performed with an analytical ultracentrifuge: at low speed, sedimentation equilibrium provides data for molecular mass determination [188, 189]; while at high speed, sedimentation velocity provides data for molecular size and shape determination [188, 189]. Analysis of TR-CP12 at a concentration of 1 mM by sedimentation equilibrium revealed a molecular mass of 14400 Da. This molecular weight is smaller than the expected 15428-15628 Da but the discrepancy can be attributed to the high concentration of the sample, as concentration is also involved in the equilibrium process (Dr Kormelia Jumel – personal communication). From this result it was concluded that the CP module pair was momoneric at the NMR concentration.

In the light-scattering technique, a beam of monochromatic light is directed through a sample and the fluctuation of the intensity of scattered light by the molecules is analyzed with a photodiode, which counts the number of photons detected as a function of time [190]. From these data the translation diffusion coefficient, $D_T$, can be extracted, which is then used to calculate the hydrodynamic radius, $R_H$, from the Stokes' equation ($R_H = k_bT /6\pi\eta D_T$, where $k_b$ is the Boltzman's constant, T is the temperature, and $\eta$ is the solvent viscosity). Finally the molecular weight is estimated from the $R_H$ [190]. Light scattering for the $GABA_B$ R1a~ TR-CP12 at the NMR concentration gave an $R_H$ value of 3.6 nm with an estimate molecular weight of 63 kDa. Such a molecular weight is consistent with a tetramer of the $GABA_B$ R1a~ TR-CP12. Interestingly, the sample has good monodispersity as judged by a polydispersity factor of only 16%.

Data from these two studies are in disagreement and it will be of interest to repeat the sedimentation equilibrium experiment on WT-CP1 as well as to perform sedimentation velocity studies. It is worth mentioning that TR-CP12 was found to run as a single peak on analytical gel filtration with a higher than expected molecular weight of about 25 kDa, although this observation

was also made for other CP modules pairs such as VCP23 (data not shown). These results are also in disagreement with the 15N T2 measurements, which suggests some degree of aggregation for TR-CP12 (section 4.1). It will be necessary to repeat the dynamic light scattering experiment as such method is prone to errors such as dust contamination.

# CHAPTER FIVE: NMR STRUCTURES OF THE TWO CONFORMERS OF THE $GABA_B$ RECEPTOR R1a COMPLEMENT MODULE TWO

The data shown in chapter four provided evidence that CP2 is a well-folded domain, and NMR experiments yielded data of good quality. Subsequently, a $^{15}N/^{13}C$-labelled sample of CP2 was produced. Standard sets of NMR experiments were collected, analysed, and used as a basis for three-dimensional structure calculations. In the course of the analysis of NMR data, it was determined that CP2 exists as two stable conformers in solution that do not inter-convert on the NMR-time scale. Consequently, a new protocol for structure calculation was devised to attempt simultaneous structure calculation of both conformers. The preliminary results of structure calculations are presented in this chapter.

## 5.1 Strategies for resonance assignment and NMR experiments utilised

Analysis of NMR spectra provides information about inter-atomic distances and torsion angles that form the basis for structure calculations. First, however, each resonance must be associated with a specific nucleus in the molecule. This process is called assignment. Amino acids have characteristic cross-peak patterns in certain 2D and 3D NMR spectra that reflect the scalar coupling of the spin 1/2 nuclei they contain. Such patterns are referred to as spin-systems. In general, resonance assignment is achieved in three sequential steps: spin-system identification, sequence-specific assignment, and assignment of the side-chain resonances.

Strategies for assignment are dependent on the size of the protein and the type of isotopic labelling. When only homonuclear and $^{15}N$ heteronuclear experiments are available, both TOCSY-, and NOESY-type experiments are needed for sequence-specific assignment. TOCSY-type spectra only provide information about spin-systems as they rely on $^{3}J$ coupling. They are analysed by matching theoretical spin-systems to experimental ones (COSY-type experiments may also be used in this part of the process). NOESY-type experiments contain cross-peaks that correlate nuclei close in space and therefore also contain the spin-system information found in the TOCSY experiments. Identified spin-systems from TOCSY experiments can be sequentially connected using NOESY-type experiments: some resonances identified from TOCSY spectra for a residue at position ($i$) in the primary sequence will exhibit cross-peaks with resonances of the residue at position ($i$+1) in the

NOESY spectra (Figure 5.1). Initially, only a limited number of amino acid spin-systems are unambiguously identified from their characteristic patterns in TOCSY-type experiments (alanine, glycine, threonine, valine, isoleucine). The remaining spin-systems are identified from a consideration of the possible identities of a particular spin-system in TOCSY spectra in the context of the possibilities for its sequence specific assignment in NOESY spectra.

When triple resonance experiments are possible, sequence specific assignment relies on sequential correlations via $^1J$ and $^2J$ heteronuclear coupling. This allows complete sequence specific assignment without knowledge of the complete spin-systems. Knowledge of the sequence-specific assignments obtained from triple resonance experiments is then used in the analysis of TOCSY-type experiments to obtain spin-system assignments.

Once spin-systems have been identified and placed sequentially, the full side-chain assignments are deduced; this is also an opportunity to assess the previous assignments.

Once all the spin-systems are fully and correctly assigned, the remainder of the NOE cross-peaks can be assigned, providing they are unambiguous. NOEs are subsequently integrated, converted into inter-atomic distance restraints, and used as input for a 3D-structure calculation. Ambiguous, unassigned, NOEs can also be of use as the relative contributions from the various components can be resolved in an iterative manner during the structure calculation process.

## 5.2 Methods used for CP2

Homonuclear and heteronuclear experiments for CP2 were collected (see table 5.1, and 5.2). All experiments were recorded at 600 MHz, 37°C, on a Varian INOVA NMR spectrometer at the University of Edinburgh, except for the $^{13}C$-edited NOESY experiment, which was recorded at 800 MHz at the University of Cambridge NMR facility. Homonuclear, and $^{15}N$ heteronuclear experiments were collected with the assistance of Dr. Dusan Uhrin and Dr. Brian Smith; triple resonance experiments were collected in collaboration with Dr. Dusan Uhrin. All samples were of 1 mM concentration in 20 mM deuterated sodium acetate, pH 4, and contained 10% $D_2O$. Spectra were processed using the AZARA software package (http://www.bio.cam.ac.uk/azara), and maximum entropy reconstruction was employed for the indirectly detected dimensions in all 3D spectra. For resonance assignment, spectra were viewed within the ANSIG (assignment of NMR spectra by interactive graphics - Per Kraulis [191]) software package. Cross-peaks were manually picked and unambiguously assigned whenever possible.

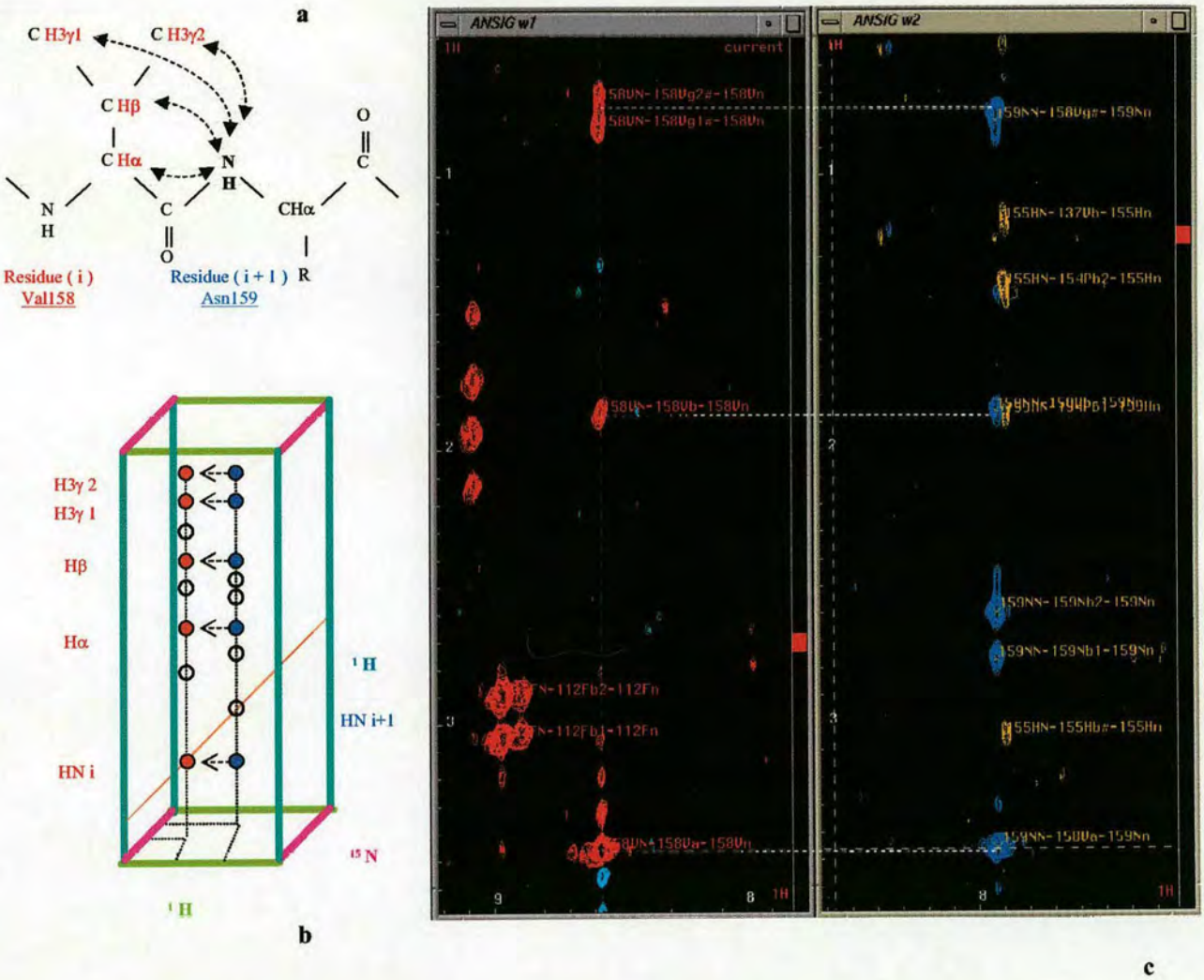



**Figure 5.1: Principle of resonance assignment in $^{15}$N-edited experiments**

a) Expected NOE connections in $^{15}$N-edited NOESY (depending on secondary structure).

The Asn159 NH "strip" may have connections with Val158 as indicated by arrows. The NHi-NHi+1 NOE may also be observed.

b) Schematic of the 3D $^{15}$N-edited TOSCY and NOESY experiments for the Val158 and Asn159 "strips".

c) View from the ANSIG setup. Left window: $^{15}$N-edited TOCSY focused on Val158 NH "strip". Right window: $^{15}$N-edited NOESY focused on Asn159 NH "strip".

## 5.3 Assigning the two conformers of CP2

The CP2 assignment was initiated using only $^{15}$N-edited TOCSY and NOESY experiments. The sequence specific assignment identified two distinct sets of resonances for 17 residues in the region of the sequence near Pro119. Dual resonances were also observed for a number of residues in vicinity of Cys99 and Cys144 (these form the first, I-III, disulphide bond). Such observations suggested the presence of two (*cis/trans*) isomers for the peptidyl-Pro119 bond. This hypothesis was consistent with data from a model of CP2 [34] in which Pro119 is predicted to be close in space to the first disulphide bond and to the N-terminus. Thus *cis/trans* isomerisation of peptidyl-Pro119 bond provided a plausible explanation for the two distinct sets of resonances. Since proline residues lack an amide proton, they are absent from $^{15}$N-edited spectra. The proline isomerisation state is in general identified from NOEs in $^{1}$H-2D NOESY. Medium to strong NOEs between proline Hδ and alpha or amide protons of the previous residue in the sequence is characteristic of a *trans* conformation. Alternatively, strong NOEs between proline Hα and alpha or amide protons from the previous residue in the sequence is characteristic of a *cis* conformation. Such discrimination requires identification of the proline spin-system. In the case of CP2, the additional resonances arising from residues affected by the *cis/trans* isomerisation process caused additional spectral overlap in the region of interest. Consequently, to allow reliable proline spin-system assignment, $^{15}$N/$^{13}$C-isotope labelling was undertaken.

Triple resonance experiments allowed full assignment of proline residues due to the additional chemical shift dispersion afforded by $^{13}$C. Backbone resonance assignment (N, HN, CO, Cα), along with complete sequence specific assignment was rapidly achieved using the combination of experiment pairs (see chapter three), 3D-HN(CA)CO/3D-HNCO and 3D-CBCA(CO)NH/3D-HNCACB. This second pair of experiments allowed assignment of the different spin systems for the two conformers of Pro119. The two distinct sets of assignments for the two conformers were manipulated by having two sets of residues in the ANSIG sequence file. One set was equivalent to the CP2 sequence numbered as in its primary sequence from Val96-Asn159 (residues from the cloning artefact were named Glu92, Ala93, Glu94, and Phe95). The other set was equivalent to the first one, but with 1000 added to every residue number, resulting in a Val1096-Asn1159 sequence. The *trans* conformer was attributed the set numbered from 96-159, while the *cis* conformer was attributed the 1096-1159 set.

**Table 5.1: NMR experiments collected on unlabelled and $^{15}$N-labelled CP2**

| Experiment | Sweep width (Hz) | Complex data points | Matrix size after processing | Processing method | Mixing time (ms) |
|---|---|---|---|---|---|
| **$^1$H** | | | | | |
| 2D COSY [130] | 10000 x 7000.35 | 4096 x 1536 | 4096 x 2048 | FFT (sinb 90, sinb 90) | |
| 2D TOCSY [192] | 8000 x 7000.35 | 2048 x 1024 | 2048 x 1024 | FFT ($sinb^2$ 90, $sinb^2$ 90) | 65 |
| 2D NOESY [135] | 8000 x 7000.35 | 2048 x 1024 | 2048 x 1024 | FFT ($sinb^2$ 90, $sinb^2$ 90) | 120 |
| **$^{15}$N** | | | | | |
| 2D HSQC [140, 141] | 5000 x 2212.6 | 1024 x 192 | 1024 x 256 | FFT ($sinb^2$ 90, $sinb^2$ 90) | |
| 2D HSQC (folded) | 5000 x 1223 | 1024 x 64 | 1024 x 128 | FFT ($sinb^2$ 90, $sinb^2$ 90) | |
| 3D $^{15}$N-edited HSQC-TOCSY | 5000 x 7000 x 1223 | 1024 x 180 x 56 | 1024 x 512 x 128 | MEM ($sinb^2$ 90) | 60.9 |
| 3D $^{15}$N-edited HSQC-NOESY [193] | 5000 x 7000 x 1223 | 1024 x 180 x 64 | 1024 x 512 x 128 | MEM ($sinb^2$ 90) | 100 |
| 2D SCT-HMSQC-HA [194] | 8000 x 2000 | 2048 x 256 | 2048 x 256 | FFT($sinb^2$ 90, $sinb^2$ 90) | |
| 3D HNHB [195] | 5000 x 6000 x 1223 | 1024 x 120 x 48 | 2048 x 256 x 128 | MEM($sinb^2$ 90) | |
| 2D HSQC in $D_2O$ | 5000 x 2015 | 1024 x 128 | 2048 x 128 | FFT ($sinb^2$ 90, $sinb^2$ 90) | |
| WEXII [196] | 8000 x 1223 | 2048 x 128 | 2048 x 128 | FFT (gauss, $sinb^2$ 75) | |
| **All experiments recorded at 37°C and 600 MHz** | | | | | |

FFT: Fast Fourier Transformation, MEM: Maximum Entropy Method. Window functions are indicated in dimension 1, dimension 2, dimension 3 order as featured in the AZARA processing scripts; for MEM, a window function was only applied in dimension 1 (acquisition dimension). $Sin^2b$ <angle> corresponds to a multiplication of the data by a sine function squared with the given angle. Gauss corresponds to a multiplication of the data by a gaussian function with a line broadening of 20 Hz, and a sharpening factor of 1.3.

**Table 5.2: NMR experiments collected on $^{15}N/^{13}C$-labelled CP2**

| Experiment | Sweep width (Hz) | Complex data points | Matrix size after processing | Processing method | Mixing time (ms) |
|---|---|---|---|---|---|
| 2D $^{13}$C-HSQC [197] | 10000 x 16000 | 2048 x 256 | 2048 x 256 | FFT (gauss, $sinb^2$ 90) | |
| 2D $^{13}$C-HSQC (folded) | 10000 x 6211.18 | 2048 x 128 | 2048 x 512 | MEM (gauss) | |
| 3D HNCO [198] | 8000 x 1780.38 x 1223.01 | 2048 x 128 x 56 | 2048 x 256 x 128 | MEM ($sinb^2$ 90) | |
| 3D HN(CA)CO [199] | 8000 x 1780.38 x 1223.01 | 2048 x 42 x 56 | 2048 x 256 x 128 | MEM ($sinb^2$ 90) | |
| 3D CBCA(CO)NH [200] | 8000 x 10559.66 x 1223.01 | 2048 x 128 x 56 | 2048 x 256 x 128 | MEM ($sinb^2$ 90) | |
| 3D HNCACB [200] | 8000 x 10559.66 x 1223.01 | 2048 x 128 x 56 | 2048 x 256 x 128 | MEM ($sinb^2$ 90) | |
| 3D HBHA(CO)NH [201] | 8000 x 10559.66 x 1223.01 | 2048 x 160 x 56 | 2048 x 512 x 128 | MEM ($sinb^2$ 90) | |
| 3D HBHANH [202] | 8000 x 10559.66 x 1223.80 | 2048 x 128 x 56 | 2048 x 256 x 128 | MEM ($sinb^2$ 90) | |
| 3D (H)C(CO)NH-TOCSY [203] | 8000 x 10559.66 x 1223.80 | 2048 x 128 x 56 | 2048 x 512 x 128 | MEM ($sinb^2$ 60) | 22.2 |
| 3D H(C)(CO)NH-TOCSY [203] | 8000 x 5599.88 x 1223.80 | 2048 x 128 x 56 | 2048 x 256 x 128 | MEM ($sinb^2$ 90) | 20.4 |
| 3D HCCH-TOCSY [204] | 8000 x 5599.88 x 4659.83 | 1024 x 160 x 64 | 1024 x 512 x 128 | MEM ($sinb^2$ 60) | 22.2 |
| 3D (H)CCH-TOCSY-CH3 [205] | 5000 x 10559.66 x 2750.08 | 1024 x 128 x 152 | 1024 x 128 x 256 | FFT ($sinb^2$ 90, $sinb^2$ 90, $sinb^2$ 90) | 20.4 |
| 3D H(C)CH-TOCSY-CH3 [205] | 5006.25 x 3600.03 x 2750.08 | 1024 x 96 x 152 | 1024 x 128 x 256 | FFT ($sinb^2$ 60, $sinb^2$ 60, $sinb^2$ 60) | 20.4 |
| 2D (HB)CB(CGCD)HD [206] | 8000 x 4000 | 2048 x 64 | 2048 x 128 | MEM ($sinb^2$ 60) | |
| 2D (HB)CB(CGCDCE)HE [206] | 8000 x 4000 | 2048 x 64 | 2048 x 128 | MEM ($sinb^2$ 65) | |
| 3D $^{13}$C-edited –NOESY [193] | 10000 x 8064.51 x 6211.18 | 2048 x 256 x 64 | 2048 x 512 x 128 | MEM ($sinb^2$ 90) | 146.5 |

**All experiments recorded at 37°C and 600 MHz – except for 3D $^{13}$C-edited –NOESY recorded at 800 MHz**

FFT: Fast Fourier Transformation, MEM: Maximum Entropy Method. Window functions are indicated in dimension 1, dimension 2, dimension 3 order as featured in the AZARA processing scripts; for MEM, a window function was only applied in dimension 1 (acquisition dimension). $Sin^2b$ <angle> corresponds to a multiplication of the data by a sine function squared with the given angle. Gauss corresponds to a multiplication of the data by a gaussian function with a line broadening of 20 Hz, and a sharpening factor of 1.3.

Characteristic, sequential, NOE cross-peaks confirmed the presence of both *cis* and *trans* peptidyl-Pro119 bonds. NOE cross-peaks were identified between Leu118 NH and Pro119 Hα in the $^{15}N$ - edited NOESY, and between Leu118 CHα and Pro119 Hα in the $^{13}C$-edited NOESY, characteristic of the *cis* conformer (Figure 5.2 - top). NOE cross-peaks were also observed between Leu118 NH and Pro119 $CH_2\delta$ in the $^{15}N$-edited NOESY, and between Leu118 CHα and Pro119 $CH_2\delta$ in the $^{13}C$-edited NOESY characteristic of the *trans* conformer (Figure 5.2 – bottom).

Side-chain resonances were assigned in two steps. First the pair of experiments, 3D-HBHA(CO)NH/ and 3D-HBHANH, was used to obtain Hα and Hβ residue shifts, then the 3D-H(C)(CO)NH-TOCSY experiment was used to achieve the remainder of the side-chain proton shifts. A similar strategy was adopted for carbon resonance assignments of the side-chains using the pair of experiments, 3D-CBCA(CO)NH/3D-HNCACB, in combination with the 3D-(H)C(CO)NH-TOCSY. Assignment of methyl resonances was also facilitated by use of the 3D-H(C)CH-TOCSY-CH3 and the 3D-(H)CCH-TOCSY-CH3.

Aromatic side-chain assignments were performed through analysis of the 2D experiment pair: (HB)CB(CGCD)HD/(HB)CB(CGCDCE)HE, which correlates Cβ of the aromatic residue to the Hδ and Hε protons respectively. Aromatic residue side-chain assignment was then confirmed in the $^{13}C$ edited NOESY.

Backbone and side-chain assignments were subsequently used to assign the 3D-HCCH-TOCSY, which was then used to assign the 3D-$^{13}C$ edited NOESY spectrum. Similarly, the 3D-$^{15}N$ edited TOCSY and NOESY were assigned as far as possible. Both 3D-$^{15}N$ edited NOESY and 3D-$^{13}C$ edited NOESY spectra contained NOEs consistent with the two distinct conformations for Pro119 mentioned above.

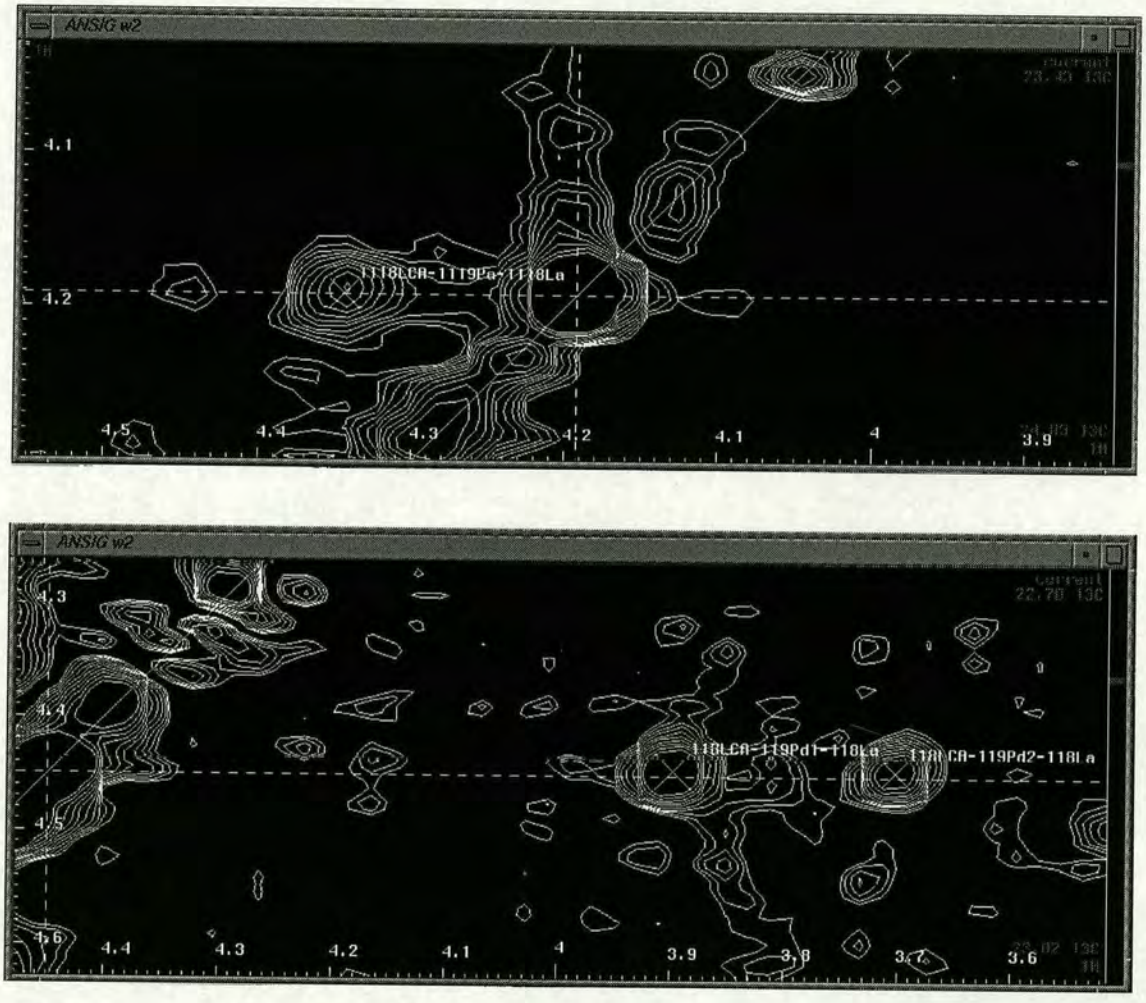



**Figure 5.2: Identification of *cis* and *trans* isoforms.**

**Top**: ANSIG view from 3D $^{13}$C-edited-NOESY experiment - strip for Leu 118 CHα in the *cis* conformer (noted 1118LCA [in F1] - 1118La [in F3]) showing NOE to Pro 119 Hα (noted 1119Pa [in F2]).

**Bottom**: ANSIG view from 3D $^{13}$C-edited-NOESY experiment - strip for Leu 118 CHα in the *trans* conformer (noted 118LCA [in F1] - 118La [in F3]) showing NOE to both Pro 119 $H_2\delta$ protons (noted 119Pd1 and 119Pd2 [in F2]).

## 5.4 Analysis of chemical shift differences between the two conformers

The $^{1}H$-$^{15}N$ 2D projection of the 3D HNCO experiment proved useful at an early stage for resolving the higher than expected number of $^{1}H$-$^{15}N$ cross-peaks, and in particular for resolving dual peaks arising from *cis/trans* conformers that were close in terms of their chemical shifts. (All of these were later confirmed by assignment of 3D spectra) (Figure 5.3, and 5.4). CP2 has 68 residues (including residues from cloning artefact), the first residue was not seen in the NMR spectra, and there are four proline residues. Therefore, 63 backbone NH peaks are expected, while 86 peaks are effectively observed, in the $^{1}H$-$^{15}N$ 2D projection of the 3D HNCO experiment. These extra backbone NH resonances were identified as belonging to 23 residues that exist in both a major and a minor conformation (see table 5.3). The *cis* and *trans* conformations of Pro119 correspond to the minor and major conformation respectively; the three other proline residues are in the *trans* conformation. A ratio of *trans* to *cis* conformer populations of 3:2 was measured from their $^{1}H$-$^{15}N$ HSQC peaks intensity ratios.

In addition to the 23, a further three residues (Leu113, Gly123, Arg125) were subsequently found to have dual NH cross-peaks but one of each pair was overlapped with other NH cross-peaks. Ile98 also yielded a doubled NH cross-peak that became observable after a period of 1-2 months. In total therefore, 27 residues were judged to have doubled ("non-degenerate") NH cross-peaks. In addition, Arg129, Cys144, Lys153 did not exhibit extra backbone NH cross-peak but exhibited dual peaks elsewhere in the spectrum (see below). Out of the 27 residues, 17 residues were found in the sequence spanning residues Val111 to Phe128.

The remainder of the 27 residues are located in two other stretches of the protein sequence that are in the vicinity of the cysteine residues involved in the first disulphide bond. The first conserved cysteine residue, Cys99, was found to exhibit a large NH chemical shift difference (both in $^{15}N$, and $^{1}H$) between the two conformations, while the third conserved cysteine, Cys144, has degenerate $^{15}N$ and $^{1}H$ resonances even though it is part of a stretch of residues that have doubled cross-peaks. This stretch also contains the conserved tryptophan residue, Trp149, and both backbone NH and side-chain NεHε resonances are non-degenerate between the two conformers. Trp149 has two distinct shifts for its Hδ1, Cζ2 and Hζ2. These atoms are located on the same side of the tryptophan side-chain. Other aromatic residues with duplication of NH cross-peaks (Phe128, Phe112) did not display doubled resonances for atoms in their side-chains.

Triple resonance-experiments allowed identification of other residues with some non-degenerate resonances between conformers: Cys144 does not have an extra backbone NH cross-peak but has a specific Hβ1 resonance for each conformer; similarly for each conformer, Arg141 has a different Hα resonance, and Gln146 has a different Hε21 resonance. A complete assignment table can be found in appendix C, and the magnitude of the chemical shift differences between conformers for the backbone atoms is shown in Figure 5.5.

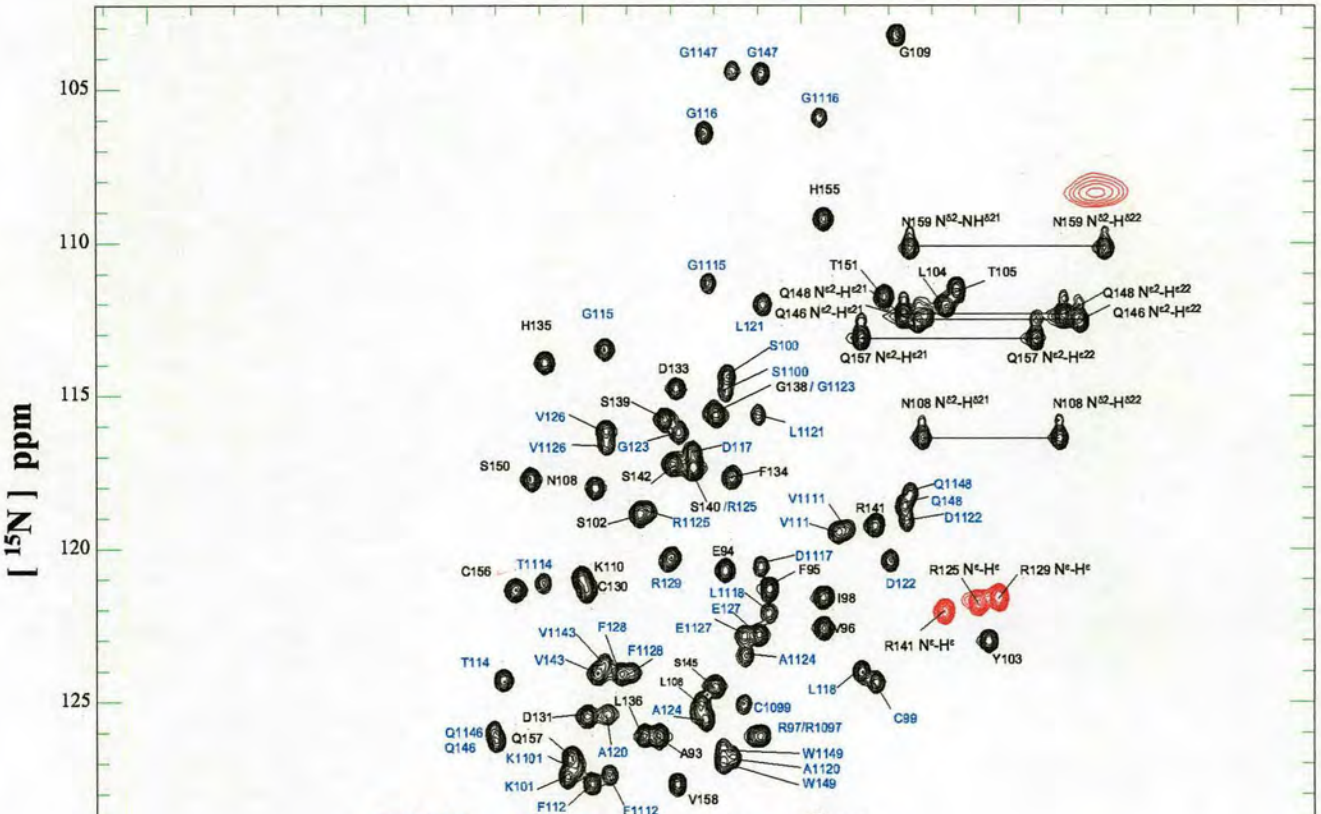



**Figure 5.3: $^{1}$H-$^{15}$N HSQC of the GABA$_B$ R1a~CP2**

Assigned residues with non-degenerate chemical shift are coloured blue for both conformers, *cis* conformer resonances were assigned with the same residue number as the *trans* conformer but one thousand was added to each of them to allow numbering distinction. Peaks of negative intensity are coloured red.

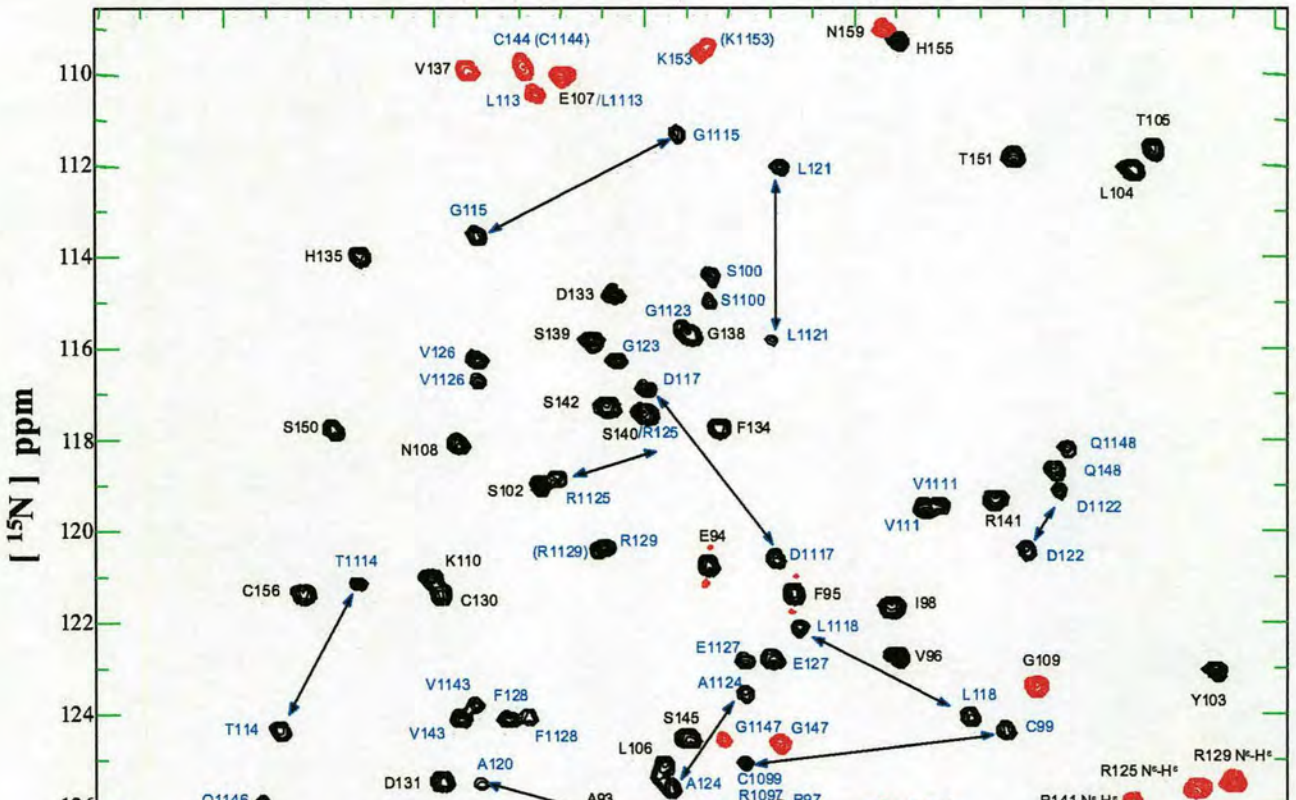



**Figure 5.4: $^{1}$H-$^{15}$N 2D projection of the 3D HNCO experiment for the GABA$_B$ R1a~CP2**

Assigned residues with non-degenerate chemical shift are coloured blue for both conformers, *cis* conformer resonances were assigned with the same residue number as the *trans* conformer but one thousand was added to each of them to allow numbering distinction. Peaks of negative intensity are coloured red. Arrows indicate large chemical shift variation between conformers.

**Table 5.3: Analysis of chemical shift differences between conformers**

| Residue | Non-degenerate chemical shifts | | | Degenerate chemical shifts in $^{13}$C-NOESY and $^{15}$N-NOESY |
|---|---|---|---|---|
| | $^{15}$N chemical shift | $^{13}$C chemical shift | $^{1}$H chemical shift | |
| Arg97 | N | | HN | complete in $^{13}$C-NOESY and $^{15}$N-NOESY |
| Ile98 | N (developed with time) | CG2, CD1 | HN (developed with time), HB,HG2#, HD1# | CAHA |
| Cys99 | N | CB | HN, HB1, HB2 | CAHA |
| Ser100 | N | | HN | CAHA, CBHB# + complete in $^{15}$N-NOESY |
| Lys101 | N | CG | HN, HB2, HGs | CAHA, CBHB1, CDHDs |
| Val111 | N | | HN | complete in $^{13}$C-NOESY + complete $^{15}$N-NOESY |
| Phe112 | N | | HN | CAHA, CBHBs and rest of side-chain in $^{13}$C-NOESY |
| Leu113 | N (overlap with Glu107) | CD1,CD2 | HN, HA, HB2, HG, HD1#,HD2# | CBHB1 |
| Thr114 | N | CA, CB, CG2 | HN, HA, HB, HG2# | |
| Gly115 | N | | HN | CAHA1 and CAHA2 |
| Gly116 | N | CA | HN, HAs | |
| Asp117 | N | CA, CB | HN, HA, HBs | |
| Leu118 | N | CA, CB, CD | HN, HA, HBs, HDs | CGHG |
| Pro119 | | CA, CB, CG, CD | HA, HBs, HGs, HDs | |
| Ala120 | N | CA, CB | HA, HB# | |
| Leu121 | N | CA, CB, CD2 | HN,HA, HBs, HD2# | CGHG and CD1HD1# |
| Asp122 | N | | HN | CAHA, 122CBHB2 and 1122CBHB# |
| Gly123 | N (overlap with Gly138) | CA | HN,HAs | |
| Ala124 | N | CA, CB | HN,HA, HB# | |
| Arg125 | N (overlap with Ser140) | CB | HN, HA,HBs, | CGHG1 CGHG2, and CDHD# + NEHE |
| Val126 | N | | HN, HA | CBHB, CG1HG1#, and CG2HG2# |
| Glu127 | N | | HN | CAHA, CBHBs, and CGHGs |
| Phe128 | N | | HN | CAHA, CBHBs and rest of side-chain in $^{13}$C-NOESY + complete in $^{15}$N-NOESY - |
| Arg141 | | | HA | complete in $^{13}$C-NOESY and $^{15}$N-NOESY |
| Val143 | N | | HN, HA, | CBHB,CG1HG1#,CG2HG2# |
| Cys144 | | CB | HB1 | complete in $^{13}$C-NOESY and $^{15}$N-NOESY |
| Gln146 | N | | HN,HE21 | complete in $^{13}$C-NOESY and $^{15}$N-NOESY |
| Gly147 | N | | HN | complete in $^{13}$C-NOESY and $^{15}$N-NOESY |
| Gln148 | N | | HN | complete in $^{13}$C-NOESY and $^{15}$N-NOESY |
| Trp149 | N, NE1 | | HN, HE1, HD1, HZ2 | complete in $^{13}$C-NOESY and $^{15}$N-NOESY |
| Total extra resonances in $^{1}$H-$^{15}$N HNCO 2D projection | 23 | | | |

A=α, B=β, G=γ, D=δ, E=ε, Z=ζ .

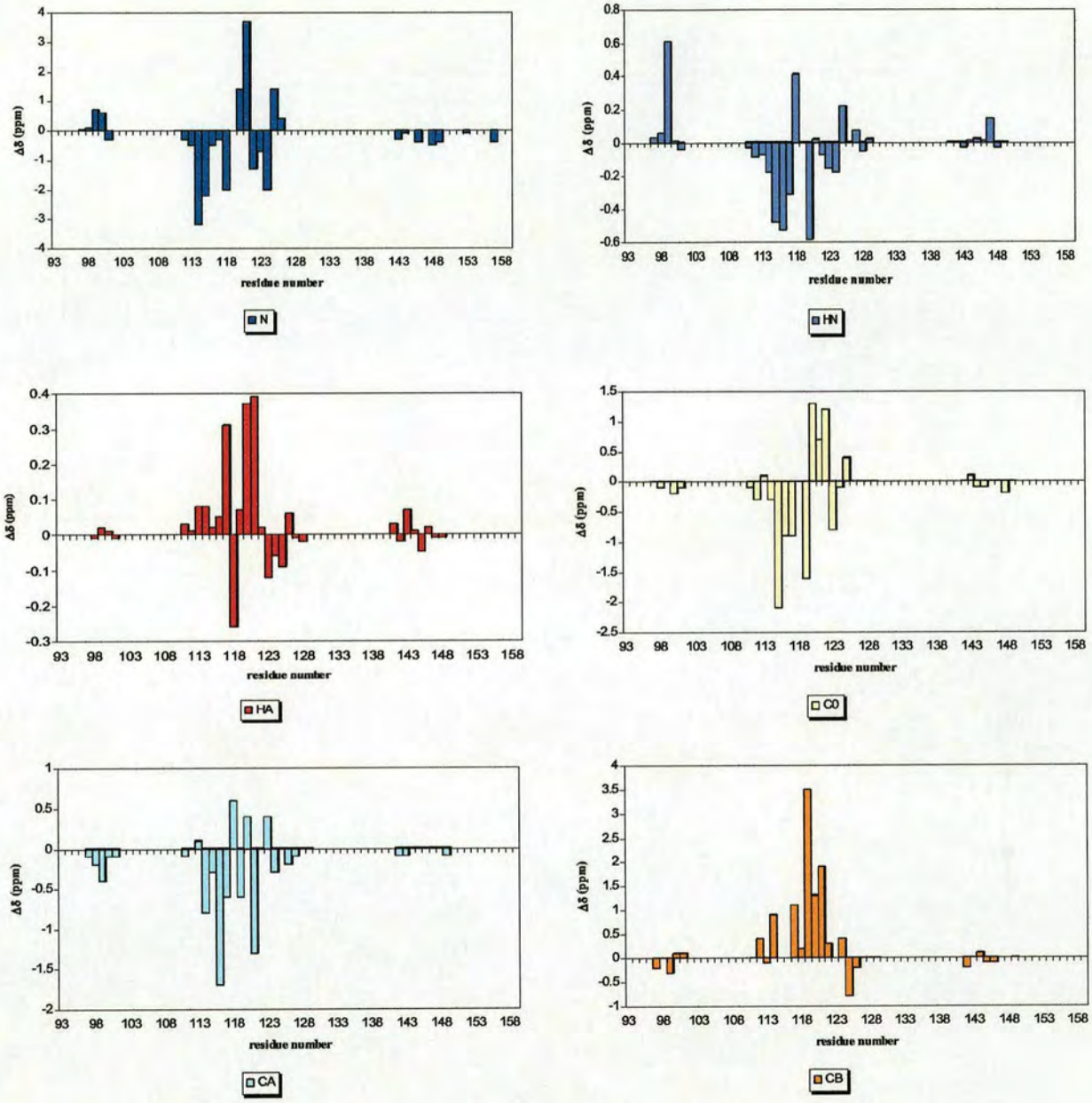



**Figure 5.5: Chemical shift differences for the two conformers (Δδ>0.01 ppm) versus residue number for the two isoforms of CP2**

Positive values represent downfield shifts and negative values upfield shifts for the minor cis-Pro119 conformer, relative to the major trans-Pro119 conformer.

## 5.5 Conversion of NOESY cross-peaks to distance restraints

NOEs are the principle source of structural data, and prior to the determination of the structure of a protein by NMR, these must be converted into relevant distance restraints. The relationship, NOE $\alpha$ $r_{ij}^{-6}$, relates NOE intensities to the distance, $r_{ij}$, between two protons, i and j. Derivation of distances from NOE volumes is, however, inaccurate in practice because local internal motions, relaxation or exchange, cross-peak overlap and possibly spin diffusion introduce significant errors. Typically, NOE intensities are classified in broad distance ranges. A strong NOE is assigned to an inter-proton distance below 2.7 Å, a medium NOE to a distance below 3.3 Å, a weak NOE to a distance below 5 Å, and a very weak NOE to a distance below 6 Å. In the current study, ANSIG software [191] was used to integrate NOE cross-peaks from $^{15}$N-edited NOESY and $^{13}$C-edited NOESY spectra. Relative volumes for each cross-peak were calculated from the normalization of the cross-peaks volumes, and subsequently categorized as strong, medium, weak, or very weak sets of inter-proton distances.

The starting point of a structure calculation is a chemical shift table that should be as complete as possible, and a partially assigned list of integrated NOE cross-peaks. These data are created from the ANSIG cross-peaks file that contains all cross-peaks from the different NMR spectra used for assignment, together with a list of integrated cross-peaks from the NOESY spectra. This file is converted into a relational database management (RDB) system format for subsequent manipulation using a set of shell scripts that are run in a sequential manner. The output is a chemical shift table of protons and their attached heavy atoms.

The *connect* program from the AZARA package is then used to match the chemical shift values of an integrated NOE cross-peak with value in the previously obtained chemical shift list within the error bound (in this work 0.03 ppm was used for $^{1}$H chemical shift and 0. 3 ppm was used for $^{15}$N/$^{13}$C chemical shifts). The *connect* program produces three files, one for unambiguous distance restraints, one for the ambiguous distance restraints, and one for the unmatched NOEs. Unambiguous and ambiguous restraints are used as input for CNS structure calculation software, while unmatched cross-peaks are inspected to detect any possible assignment problems or peak-picking errors.

## 5.6 Structure calculation protocol

### 5.6.1 General principle

NOEs provide the primary experimental information on the local spatial relationship of protons. The positions of other atoms have to be derived from protein chemistry; this *a priori* information is conveniently expressed as energy functions within a molecular dynamics force field. Finding conformations of the molecule that satisfy both the experimental data and the force field describing the covalent structure and non-bonded interactions is equivalent to finding conformations that minimize a hybrid energy function ($E_{hybrid}$), which has contributions from both experimental data and the force field: $E_{hybrid} = E_{chem} + \omega_{nmr} E_{nmr}$ where $E_{CHEM}$ describes empirical chemical information. $E_{nmr}$ represents the NMR data by describing the difference between observed and calculated data, and $\omega_{NMR}$ is its weighting factor. The empirical energy is a sum of energy terms: $E_{chem} = \omega_{cov} E_{cov} + \omega_{vdw} E_{Vvdw}$ where $E_{cov}$ describes covalent geometry (bond lengths, bond angles, and planarity) and $\omega_{COV}$ is its weighting factor. $E_{Vvdw}$ is the van der Waals function and corresponds to a purely repulsive function; $\omega_{vdw}$ is its weighting factor. Electrostatic interactions and van der Waals attractive functions are usually only included in structure calculations at the refinement stage. A common form of $E_{nmr}$ describes NOE derived distance restraints and dihedral angle restraints derived from J-coupling constants: $\omega_{NMR} E_{nmr} = \omega_{noe} E_{noe} + \omega_{dihedral} E_{dihedral}$.

The CNS software was used to calculate 3D structures from randomised coordinates using simulated annealing with Cartesian dynamics under NOE-derived distance restraints. This type of optimisation of the hybrid energy function uses molecular dynamics based on simulated annealing [207]. Molecular dynamics (MD) consists of simulating the motion of a system of particles, here the atoms of the molecule. MD is equivalent to solving Newton's equations of motion as a function of time by computer simulation techniques: $a_i = \frac{F_i}{m_i} = -\nabla_i E_{hybrid} = m_i \frac{d^2 r_i}{dt^2}$, where $a_i$ is the acceleration; $F_i$ the force exerted; $m_i$ is the mass, and $r_i$ is the coordinates of atom i. Atoms are given initial velocities taken at random for a low temperature, then simulation is performed for a few picoseconds; the sum of the forces exerted on each atom induces a change in velocity, which is used to find their new positions and their associated $E_{hybrid}$. In this type of simulation, the temperature is held constant and atomic velocities are adjusted to remove excessive kinetic energy. The potential energy of the system as a function of the atomic coordinates defines an energy surface, whose wells represent the

different possible stable or metastable structures of the molecule. Molecular dynamics can overcome local energy barriers in $E_{hybrid}$ but large energy barriers can only be crossed at high temperature. In simulated annealing (SA), the system is heated at high temperature (up to 2000 K) followed by a slow cooling period; and this process is repeated to obtain the global energy minimum. The independent scaling of each contribution in $E_{hybrid}$ gives rise to a large number of possible SA schemes. But generally the bad geometry of the initial structure dictates that covalent geometry terms must start with very low weights, and these are gradually increased during the calculation. Convergence is achieved through a multi-stage protocol starting from random Cartesian coordinates, followed by several steps of SA and MD, and other mathematical energy minimization methods (Powell, and conjugate gradient minimisation). A script referred to as the "wrapper" script was used to drive the sequence of the calculation and to invoke appropriate restraint files (Raine, Smith, and Domaille-unpublished, see appendix D-1). Each round of calculations includes three steps; randomisation (*rand.cns* – see appendix D-2), regularisation (*rrsa_swap.cns* – see appendix D-3) and a final refinement step (*refine_swap.cns* – see appendix D-4). These steps correspond to three versions of the XPLOR protocols *rand.inp*, *rrsa.inp (DGSA.inp)* and *refine.inp* that have been modified to allow floating stereochemistry and include active swapping of pro-chiral centers [208] (methyl, methylene, side-chain amide). Only floating assignment assures that the assignment of such groups is consistent in the structure. Approximate treatment of pro-chiral groups with pseudo-atoms or sum averaging leads to a loss of information, i.e. the correlation of NOEs to the single protons or methyl group of the pro-chiral group is lost.

The ARIA (Ambiguous Restraints for Iterative Assignment) suite of scripts [209] within CNS is often employed for iterative automated NOE assignment. In this study, a more conservative approach was used [210], and no iterative automated NOE assignments were performed. ARIA was utilised only as an analytic tool to select ambiguous NOE contributions from calculated structures that showed convergence.

An ambiguous NOE corresponds to a sum of potential individual contributions. It can be used to construct an ambiguous distance restraint that is added to the restraint lists and used for structure calculation. Then $E_{noe}$ is calculated by defining an effective distance; this is a "summed distance" with contributions from more than one proton pair [211].

The XPLOR/CNS averaging option 'sum' was used to impose ambiguous distance restraints in the calculation according to the following equation: $\overline{D} \equiv \left( 1/N_{monomer} \sum_{a=1}^{N\delta} d_a^{-6} \right)^{-1/6}$ where a runs through

all $N\delta$ contributions to a cross-peak, $d_a$ is the distance between two protons corresponding to the $a^{th}$ contribution, and $N_{monomer}$ is use to scale the distances corresponding to ambiguous NOE cross-peaks in symmetric multimers. The 'sum' averaging option (above) is generally preferred to the ensemble average option ($<d^{-6}>^{-1/6}$) as it produces an effective distance, which is closer to the desired result for an ambiguous NOE cross-peaks [212]. The difference between sum average and ensemble average is that in the sum average option, each contribution is treated separately and summed to give an effective distance rather than been an average.

A first round of calculation is generally performed in which 20 structures are calculated using NOE-derived restraints and the three-step protocol described above. Structures are normally analysed in terms of energy statistics, such that structures with the lowest NOE energies are selected and analysed for NOE violations (>0.5 Å) using the ARIA based script, *analyse* (see appendix E-1). The calculated structures are also analysed to monitor convergence. This first iteration is repeated until all NOE violations are resolved (by re-examination of the spectra and the cross-peak lists), then selected structures with lowest NOE energy are used to filter the original ambiguous NOE restraints with the ARIA based script [209], *filter* (see appendix E-2). Small contributions to ambiguous NOEs are discarded, and a new iteration is run to generate new structures. Duplicate NOE restraints are also discarded with the ARIA based script [209], *check* (see appendix E-3). This process is repeated until structures with lowest NOE energy and data do not change significantly. To assess convergence the root-mean-square difference (RMSD) between corresponding atoms amongst the ensemble of structures is computed [207]. The UWMN (Hartson and Caves - University of York) program was used to calculated RMSDs for structure ensembles; in particular the RMSDs of the backbone atoms (Cα, N, and CO) from each structure in the ensemble relative to the mean structure are often used to assess convergence.

### 5.6.2 Protocol for calculation of the two conformers and preliminary structures

In the case of CP2, proline isomerisation seems to induce only localized perturbations. Therefore (as described earlier) while many cross-peaks are doubled, most cross-peaks are degenerate and can arise from either of the two conformers. It is clearly impossible to distinguish the exact contribution of each conformer to the degenerate NOE peak. Different strategies have been reported to deal with *cis-trans* peptidyl-proline conformers. One of them is simply to avoid the problem, and use site-directed mutagenesis [213, 214] to exchange the proline residue for a glycine. However, such an approach is only sensible if the mutant protein remains functional, and the structure is known not to be perturbed in any way. In the case of CP2 both biological function, and the configuration of the peptidyl-proline bond in the intact protein, are unknown - therefore the determination of the structures of both conformers is preferable. Another approach is to calculate separate structures for the conformers, allowing degenerate NOE cross-peaks to be employed in both conformer calculations. Interestingly, it has been shown that when NOEs are equally weighted for both conformers, the resulting structures do not satisfy the entire set of restraints [215]. Nevertheless, Yuan et al [216] showed that this type of calculation was possible when an appropriate calibration of the NOE intensity is used to compensate for the differences in populations of the two forms.

Proline isomerisation also generates an increased ambiguity when assigning NOE data. Manually resolving such ambiguities is laborious. In this study a system where the structures of both *cis* and *trans* conformers are calculated simultaneously was devised to allow testing of ambiguous NOE against *cis* and *trans* conformers. This approach does not solve the respective *cis/trans* contribution problem to a NOE, but scaling NOE data according to their respective conformer (see above) would require extensive manual assignment. In this study, a restrained non-crystallographic symmetry (ncs) term was used to introduce an equal contribution to a NOE from each conformer. This represents a reasonable first approximation of the 3:2 *trans* to *cis* conformer ratio population described in section 5.4. This was done in collaboration with Dr. Brian Smith (Institute of Cell and Molecular Biology - Edinburgh University).

All NOE peaks were treated initially as ambiguous, so that the resolved doubled peaks will be correctly assigned to the appropriate conformer by the *connect* program, while degenerate peaks will be ambiguously assigned to either conformer (a distinction must be made here between "ambiguous" assignments that arise from resonances within the same conformer, and those that are ambiguous because they have essentially the same chemical shift in both conformers). The conformers were

declared in CNS as two monomers in the same calculation bath (two segids were declared, as well as a monomer statement of two), and symmetry between the two conformers was maintained by use of a restrained non-crystallographic symmetry (ncs) term. A least-square superposition of the *cis* ncs-equivalent specified atoms onto the *trans* ncs-equivalent specified atoms is computed, and the average x,y,z for each atoms is taken. Each atom is then restrained according to the $E_{ncs}$ energy term, $E_{ncs} = \omega_{ncs} (x-\langle x \rangle)^2$, where $\omega_{ncs}$ is used to weight the restraint. This was necessary because of a limitation in the sum averaging treatment of NOEs- once a restraint is satisfied in one of the two conformers it does not have to be satisfied in the other one.

The protocol also restricted Pro119 conformation to be *cis* for the minor form and *trans* for the major form. The *rand.inp* script was modified to create identical random initial coordinates for the two conformers, which were then pulled apart by use of a translation. A chemical shift list was built from assigned cross-peaks as described previously, and the chemical shifts for degenerate peaks were allocated to both *cis* and *trans* conformers. This was achieved by writing an awk routine, *fillShifts* (see appendix F-1) that runs along with the set of shell scripts described in section 5.5. The NOE cross-peaks file obtained from ANSIG was stripped of all assignment using a script, written for this purpose; *stripAssignments* (see appendix F-2). Subsequently, the modified chemical shift list and the NOE cross-peaks file stripped of all assignments were used as input of *connect*. The *connect* output restraint file was further modified by an awk routine, *rdb2orrest.awk*, to express ambiguous restraint in an *or-restraints* format (see appendix F-3). This format allows ambiguous restraints to be ambiguous only between atoms that are present in the same conformer (see appendix G).

Each round of calculations produced 20 calculated structures. Five preliminary rounds of calculations were necessary to adequately separate the two monomers in the first step of the protocol, and to identify incorrect assignments of one Phe residue side-chain. Four to five heavily overlapped peaks including some close to the water signal in the $^{15}$N-edited NOESY, were discarded. *Cis* conformer structures were poorly defined in the first attempt that did not included the ncs term. This symmetry term was subsequently applied to the full molecule (ncs between residues 92-159 and residues 1092-1159, mentioned later as "*full* ncs"). From the last of these preliminary rounds of calculation, ten of the lowest NOE energy structures exhibited the CP module-fold, and were without any significant NOE violations; these were selected for the first "filtering" step. The *filter* script was used to eliminate assignment possibilities that contributed less than 1% to an ambiguous NOE. The output was a new restraint file, which was subsequently used in a new round of structure calculation.

Eight of the lowest NOE energy structures for each conformer along with their energy statistics are presented in Figure 5.6 a, and table 5.4.

Using the same filtered set of restraints, a new calculation round was performed in which the ncs term was maintained only for residues exhibiting degenerate chemical shifts (ncs between residues 92-96, 102-110, 129-140, 142, 145, 150-159 and residues 1092-1096, 1102-1110, 1129-1140, 1142, 1145, 1150-1159 respectively). This results in "twisted" structures for the *cis* conformer, where the hypervariable loop (residue 112-120) points towards the core of the module, or in structures having poorly define region towards the first disulphide bond, despite a lack of NOE violations. Such an observation was consistent with the analysis of the chemical shift differences between conformers as presented in table 5.3- only a few "extra" (*cis* conformer-derived) peaks are observed from those residues in the region of the first disulphide bond, and this does not provide enough information to fully define this region in the *cis* conformer. Subsequently, a calculation where the ncs term was maintained for the whole molecule, with the exception of residues 111 to 128, was performed (ncs between residues 92-111, 128-159 and residues 1092-1111, 1128-1159 respectively, mentioned later as "*reduced* ncs"). These residues, which encompass the "hypervariable" loop exhibit most of the double resonances in the molecule. At this stage, 75 $^{3}J_{(HN\text{-}H\alpha)}$ derived dihedral angle restraints derived from semi-constant-time $^{1}H$-$^{15}N$ –HMSQC-HA experiments [194] for both conformers were added into the calculation. In the resulting structures, no NOE violations were observed; 20 of the 75 $^{3}J_{(HN\text{-}H\alpha)}$ values were violated but not by more than one Hz. Eight of the lowest NOE energy structures for each conformer along with their energy statistics are presented in Figure 5.6 b, and table 5.5.

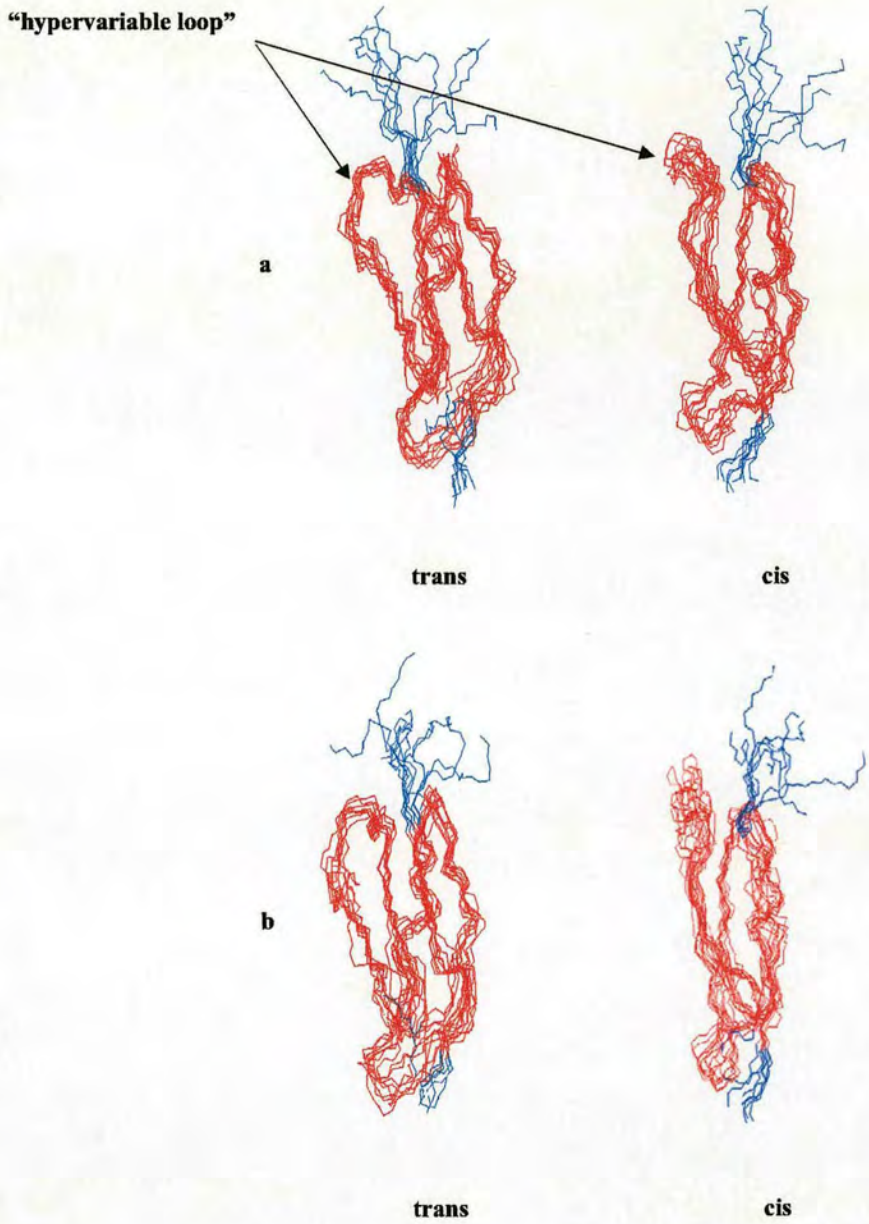



**Figure 5.6: Comparison of calculated structures with *full* or *reduced* ncs**

**(Views from molmol program [http://www.mol.biol.ethz.ch/wuthrich/software/molmol])**

a) Overlay of the eight lowest NOE energy structures calculated with *full* ncs.

b) Overlay of the eight lowest NOE energy structures calculated with *reduced* ncs.

Residues from Cys99 to Cys156 are coloured red.

**Table 5.4: Energy statistics for 8 lowest NOE energy structures from round with *full ncs* (ncs between residues 92-159 and residues 1092-1159).**

| Structure (both conformers) | $E_{total}$ | $E_{bonds}$ | $E_{angle}$ | $E_{improper}$ | $E_{vdw}$ | $E_{noe}$ | $E_{dihedral}$ | $E_{ncs}$ |
|---|---|---|---|---|---|---|---|---|
| #1 | 601.48 | 4.41 | 57.33 | 5.43 | 103.33 | 10.60 | 236.24 | 184.11 |
| #2 | 626.65 | 4.93 | 61.59 | 6.01 | 113.35 | 13.77 | 262.74 | 164.23 |
| #3 | 653.54 | 5.21 | 66.47 | 5.15 | 126.89 | 16.34 | 242.17 | 191.29 |
| #4 | 691.89 | 5.70 | 65.54 | 6.15 | 119.16 | 23.64 | 250.83 | 220.84 |
| #5 | 578.01 | 4.98 | 60.73 | 5.85 | 122.33 | 24.07 | 217.31 | 142.71 |
| #6 | 617.70 | 5.30 | 64.42 | 7.03 | 121.75 | 24.19 | 242.61 | 152.37 |
| #7 | 626.05 | 4.78 | 59.28 | 6.18 | 111.05 | 30.17 | 259.86 | 154.70 |
| #8 | 595.85 | 5.53 | 61.40 | 4.76 | 117.67 | 31.23 | 213.16 | 162.07 |

Overall backbone RMSD: Cys99-Cys156 = 0.98 Å / Cys1099-Cys1156 = 1.06Å

**Table 5.5: Energy statistics for 8 lowest NOE energy structures from round with *reduced ncs* (ncs between residues 92-111, 128-159 and residues 1092-1111, 1128-1159 respectively).**

| Structure (both conformers) | $E_{total}$ | $E_{bonds}$ | $E_{angle}$ | $E_{improper}$ | $E_{vdw}$ | $E_{noe}$ | $E_{coupling}$ | $E_{dihedral}$ | $E_{ncs}$ |
|---|---|---|---|---|---|---|---|---|---|
| #1 | 680.61 | 4.58 | 66.31 | 5.90 | 123.77 | 12.86 | 14.77 | 304.20 | 148.18 |
| #2 | 724.24 | 6.25 | 75.45 | 9.32 | 127.38 | 15.72 | 16.97 | 308.92 | 164.20 |
| #3 | 642.58 | 5.88 | 73.04 | 7.76 | 126.65 | 16.06 | 17.53 | 258.27 | 137.36 |
| #4 | 665.24 | 6.52 | 72.07 | 5.95 | 134.83 | 16.52 | 20.07 | 279.07 | 130.17 |
| #5 | 679.92 | 6.34 | 70.56 | 6.24 | 145 | 17.1 | 33.43 | 276.79 | 124.42 |
| #6 | 662.66 | 5.89 | 73.62 | 8.48 | 119.44 | 21.84 | 18.32 | 271.50 | 143.53 |
| #7 | 877.53 | 8.85 | 105.92 | 14.09 | 162.39 | 21.94 | 23.21 | 314.47 | 226.62 |
| #8 | 825.51 | 6.42 | 80.08 | 9.81 | 126.51 | 26.24 | 26.98 | 324.59 | 224.85 |

Overall backbone RMSD: Cys99-Cys156 = 1.16 Å / Cys1099-Cys1156 = 1.4 Å

**Table 5.6: Restraint analysis**

| Unambiguous restraint after filtering | *trans* | *cis* |
|---|---|---|
| intra-residue | 128 | 124 |
| sequential | 62 | 42 |
| short range ($2 \leq \|i-j\| \leq 4$) | 13 | 5 |
| long range ($\|i-j\| > 4$) | 68 | 21 |
| Total | 271 | 192 |

Ambiguous NOE left: 2219 (total) –271-192=1756

## 5.7 Analysis of the preliminary structures

At this stage of the calculation, average backbone RMSDs (based on Cα, N, and CO atoms of residues Cys99-Cys156 for each conformer) for the 8 best structures was around 1 Å in each conformer when full ncs is applied; while average backbone RMSD values for calculation with reduced ncs are somewhat higher (1-1.4 Å). However, the *cis* conformer structures calculated without ncs for residues Val111-Phe128 [residues with the most non-degenerate (between conformers) chemical shifts] show some differences in this region when compared to structures calculated with ncs applied to the full molecule (Figure 5.7 a). Even though structures calculated with reduced ncs are less precise, they are surely more accurate. It is therefore probably preferable to perform reduced ncs type of calculations, and use NOE-derived restraints as well as dihedral angle restraints to define this region of the molecule (14 *cis*-specific dihedral angle restraints are available in the sequence portion spanning from Val111 to Phe128). Such average backbone RMSD values along with low NOE energy ($E_{noe}$) show that the structures of each conformer are converging towards a fold that satisfies experimental NOE restraints. Both of the conformers calculated using the reduced ncs term are (unsurprisingly) relatively similar with the exception of the non-restrained (by symmetry) hypervariable loop region (Figure 5.6, 5.7 b, and 5.8). The expected secondary structure elements predicted from sequence comparison, $^3J_{(HN\text{-}H\alpha)}$ values and consensus chemical shift index are found in the preliminary structures of both conformers (Figure 5.9). In particular both conformers have the helical turn inferred from the consensus chemical shift index [217], although not all of the beta-strands are found using the Kabsch and Sander criteria [218] (see Molscript representation [219], Figure 5.10 a, b, and c).

At this stage of the calculation, structures have high dihedral and ncs energies ($E_{dihedral}$, and $E_{ncs}$ respectively). However high $E_{dihedral}$ values are to be expected, as such values do not correspond to one but two molecules. Also the PARALLHDG-5.1/2 forcefield used in this study introduce dihedral angle potentials in the regularisation and refinement stages which improve $\phi$, $\psi$, and $\chi_1$ angles at a cost in $E_{dihedral}$. Finally, it is worth noting that the $E_{ncs}$ observed for CP2 *cis-trans* calculations are four times lower than the ones observed in the case of the mouse heterochromatin 1 dimer using the same weighting factor [220] (weighting factor is 0.05).

Future rounds of structure calculation will include hydrogen-bond information, which will certainly increase secondary structure element definition. A series of $^{15}N$-HSQC spectra acquired on a $^{15}N$ labelled of CP2 freshly dissolved in $D_2O$ have identified a number of slowly exchanging amides likely to be involved in hydrogen-bonds (Figure 5.11). Some of these residues have also been

identified from 2D heteronuclear water exchange filter (WEXII-FHSQC) experiments [196]: these are Leu104, Val126 (*trans*), Glu127 (*trans*), Asp131, Cys144, Gln148 (*cis* and *trans*), and Gln157. Also to be incorporated are $^3J_{(H\alpha\text{-}H\beta)}$ values measured from the 3D HNHB experiment [195], which relate to $\chi_1$ angles (data not shown).

Taken together the results obtained to date confirm the expected CP-module fold for both conformers. On the basis of sequence alignment Pro119 was predicted to be located as the last residue in the hypervariable loop. In the preliminary structures, Pro119 is indeed located at the end of the hypervariable loop. Such a location would mean that *cis-trans* peptidyl-Pro119 isomerisation could occur with minimum conformational adjustment of the core of the protein.

After filtering, a total of 2219 NOE was used in the calculations, from which 1756 were ambiguous, while 271 were unambiguous and specific to the *trans* conformer and 192 were unambiguous and specific to the *cis* conformer (see table 5.6). Although, at the present stage of the project, the NMR-derived structures of both conformers are converging, and satisfy NOE restraints, it will be necessary to assess more accurately NOE assignments made by the filtering process as these statistics do not take into account the number of unambiguous NOEs classified as ambiguous simply because these are applied to both conformers.

As an alternative approach to the novel one described above, it will be of interest to compute conformer structures according to Yuan's protocol. This will provide an instructive comparison with the current one.

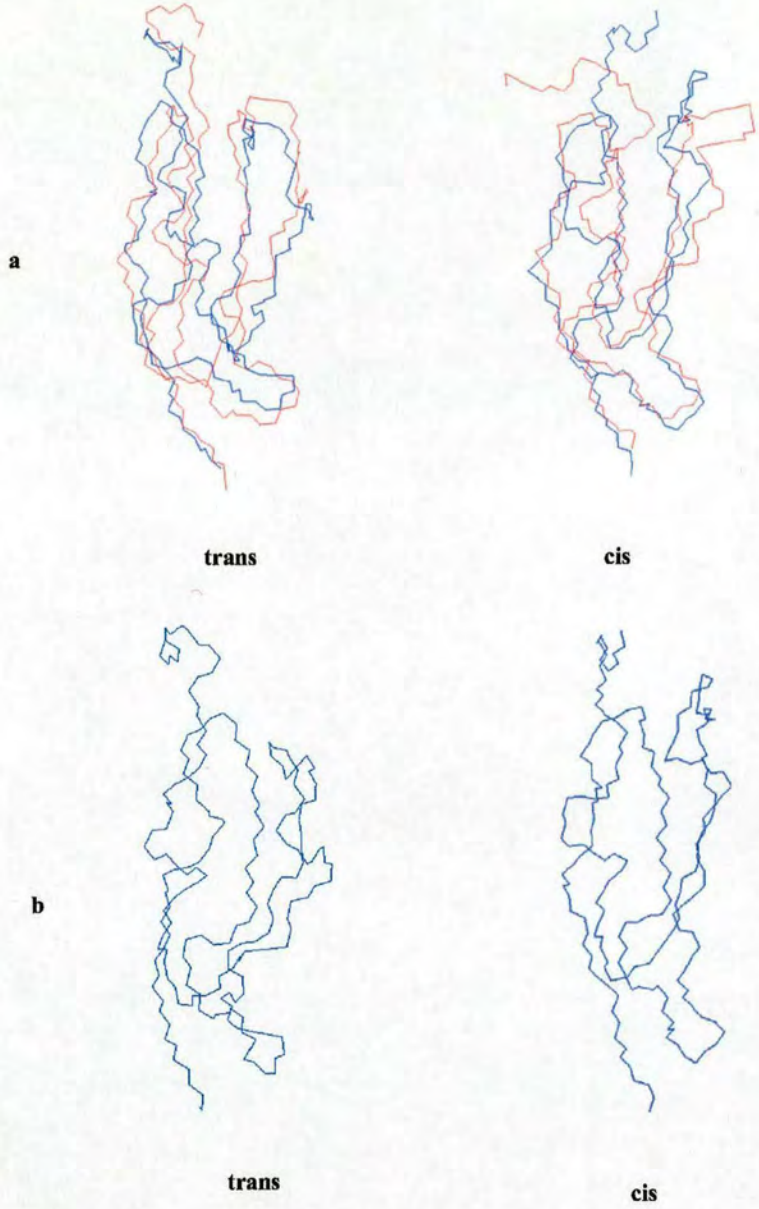



**Figure 5.7: Comparison of calculated structures with *full* and *reduced* ncs (views from molmol program)**

a) Overlay of closest to mean structures from the ensemble of eight lowest NOE energy structures calculated with *full* (red) or *reduced* (blue) ncs.

b) Closest to mean structure from the ensemble of eight lowest NOE energy structures calculated with *reduced* ncs.

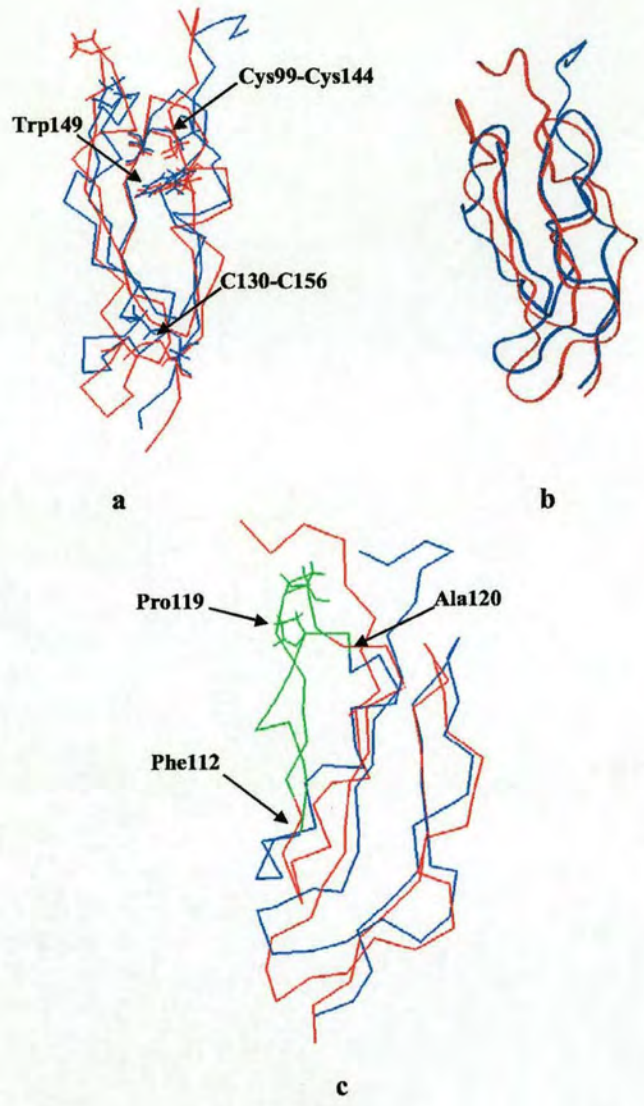


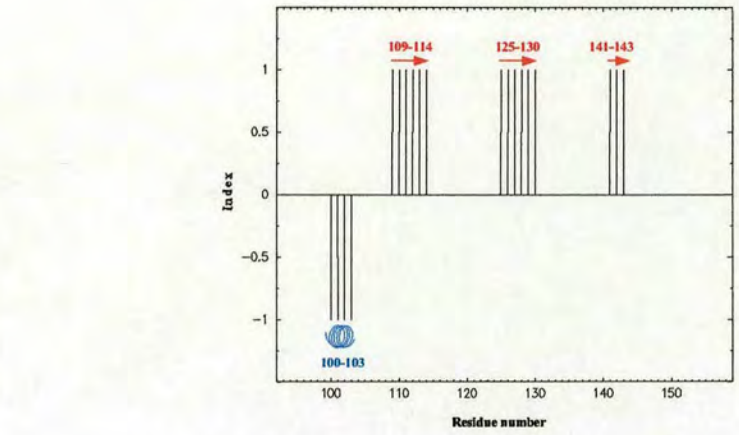

**Figure 5.8: Comparison of the *cis* and *trans* conformers of CP2**

**(Views from molmol program and InsightII program [BIOSYM])**

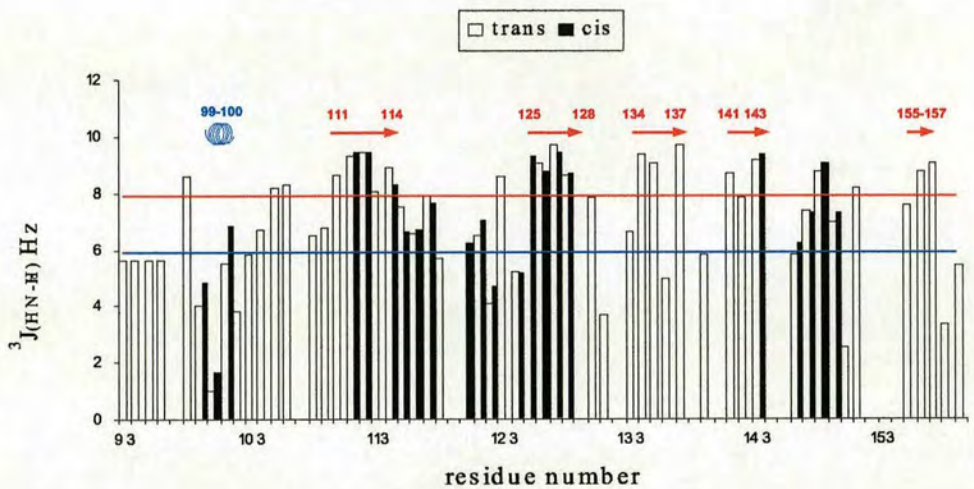
a) Overlay of closest to mean structures from the ensemble of eight lowest NOE energy structures calculated with *reduced* ncs - red: *cis* conformer, blue: *trans* conformer.

b) Same as (a)- ribbon plot.

c) Same as (a)-the "hypervariable loop" predicted from sequence alignment is coloured green.




a




b

**Figure 5.9: Secondary structure elements observed and predicted in the *cis* and *trans* conformer of CP2**

a) The consensus chemical shift index of CP2.

The consensus chemical shift index was calculated using $^1H\alpha$, $^{13}C\alpha$, $^{13}C\beta$, and $^{13}CO$ chemical shifts, in the CSI program obtained from the university of Alberta (Canada) [217]; results are identical for both conformers. The observed secondary structure is illustrated by a spiral symbol for helix and arrow symbols for β-strands.

b) Comparison of the $^3J(HN\text{-}H\alpha)$ values for the *cis* and *trans* conformers, β-strand are predicted from values > 8Hz, α-helices are predicted from values < 6 Hz [207].

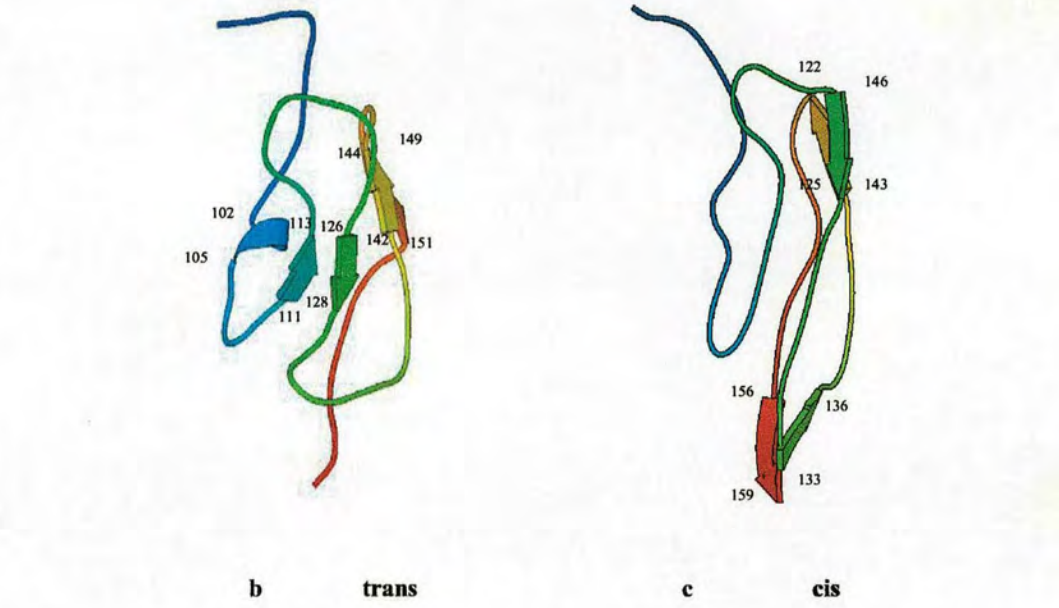



**Figure 5.10: Secondary structure elements observed and predicted in the *cis* and *trans* conformer of CP2**

a) Primary sequence of CP2 and consensus secondary structure elements derived from the consensus chemical shift and the $^3J_{(HN-H\alpha)}$ values. β-strands are coloured red, and the α-helix is coloured cyan.

b) Molscript representation for the *trans* conformer (closest to mean structure)-reduced ncs.

c) Molscript representation for the *cis* conformer (closest to mean structure)-reduced ncs.

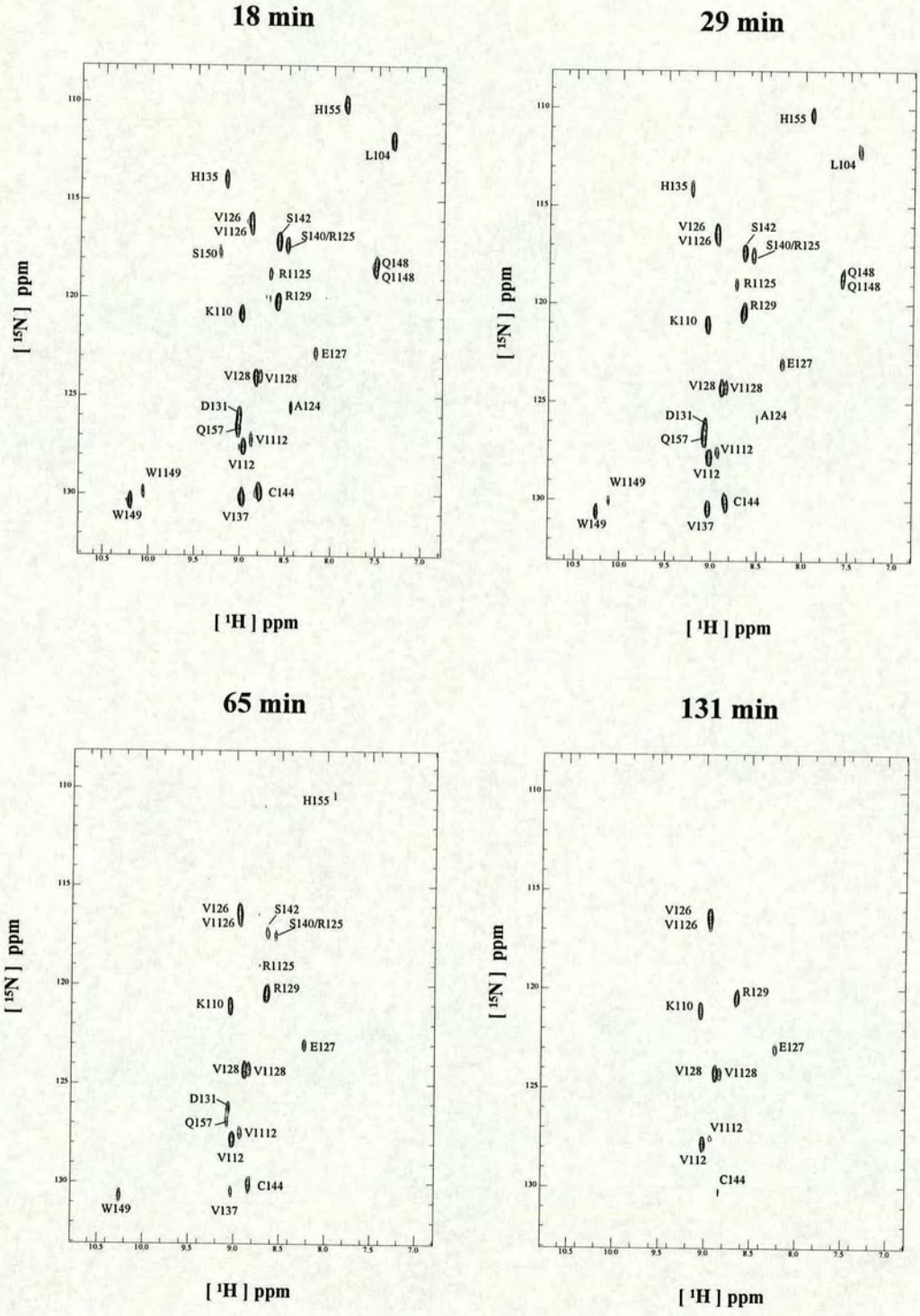



**Figure 5.11: $D_2O$ exchange $^{1}H$-$^{15}N$ HSQC experiments**

Sample was freeze-dried and redissolved in $D_2O$; time indicates elapsed time after $D_2O$ addition to the protein sample.

# CHAPTER SIX: DISCUSSION

The intentions of this short chapter are: to summarise the main findings of the various components of this thesis; to draw them together so as to provide overall conclusions; and to consider briefly the outcome in the context of the literature, and of the wider field of $GABA_B$ R1a structure and function. This last aim inevitable involves some rather speculative discussion.

The basis of this thesis is that the CP modules found in the $GABA_B$ receptor are biologically significant. No function was assigned to these modules when they were first discovered. Four years later, their role still remains a mystery. In this respect, it is worth pointing out that amongst the relatively small number of known $GABA_B$ receptor variants, three involve the CP modules (R1a [13], R1c (unpublished –Genbank accession number AJ012187) and R1e [50]). Recently a new subtype, $GABA_B$ R1g, has been cloned [221], it consists of the sequence of both CP-modules from R1a subunit, and an additional 80 residues. *In vivo* expression of this soluble protein will need to be confirmed, as there has been no further reports in the literature of such a variant; but if real, it provides further evidence for a role of CP modules. So what has been learnt about the CP modules (and their potential ligand(s)) from the current work?

## *The first CP module of $GABA_B$ R1a*

The sequence of $GABA_B$ R1a corresponding to the first putative CP module (CP1) was expressed in useful quantities (8-10 mg per litre of pure protein) in the methylotophic yeast *P. pastoris*. Thus CP1 joins a list of CP modules expressed successfully in this system – membrane cofactor protein (CD46) module 1 (Dr. Joanne O'Leary – PhD thesis), VCP (intact and in various fragments) [222], CR1~1517 (with N-glycosylation sites ablated) [210], DAF (in various fragments of two, three and four modules, all of them His-tagged) [223]. Other module-types have also been produced at useful yields in *P. pastoris*; examples include EGF-like modules [224] and CUB modules [225]. In contrast, attempts at expressing $GABA_B$ R1a CP1 in insect cells and *E. coli* have not so far proved to be fruitful (data not shown). The yield from insect cells has proved to be very low so far. *E. coli*-expressed material is insoluble (cytosolic expression - data not shown) or non-detectable (periplasmic expression – not shown). While re-folding of insoluble *E. coli*-expressed CP modules has been reported (in the case of one- and three-CP fragments from CR1) [226] it appears that there is no generally applicable

re-folding protocol for such proteins and time did not allow a full exploration of this kind of strategy. Thus it appears that *P. pastoris,* our original choice of expression vector, is the most consistently reliable and efficient one for CP1.

The recombinant CP1 was characterised by mass spectrometry which confirmed the expected mass, and further studies using mass spectrometery and other techniques confirmed the expected (cysteines I-III, II-IV) disulphide pairing. Biophysical characterisation revealed that *P. pastoris*-expressed CP1 is highly soluble (~ 1 mM) and stable, but does not adopt a well-defined structure. More specifically, NMR data were suggestive of the presence of multiple conformers of CP1 in intermediate exchange, while no transition was observed in DSC. Far-UV CD was consistent with a secondary structure-content (high in β-strands) similar to that of other CPs, but when collected as a function of temperature the far-UV CD spectra indicated a lack of a cooperative-unfolding event. Near-UV CD and NMR NOESY spectra, however, indicated the presence of some tertiary structure (i.e. buried aromatic residues) in the protein. The fact that the I-III and II-IV disulphides have been formed in CP1 is also consistent with some tertiary structure formation. Thus it appears that *P. pastoris* CP1, expressed in isolation from the remainder of the receptor, has some characteristics of a molten globule (it also binds ANS) but its conformational flexibility is restricted by its disulphide bonds and hydrophobic interactions between side-chains. (The status of CP1 as a "molten globule" was discussed at some length in chapter 4).

The crucial question (and one that will be considered further below) clearly then arises – is this recombinant form of CP1 representative of the conformation of CP1 in the context of the intact native $GABA_B$ receptor? Not only does one need to consider the truncated nature of the expressed fragment but it should be remembered that the receptor is glycosylated and contains a second subunit – $GABA_B$ R2. Unfortunately, since none of the other expression systems tried have yielded useful amounts of protein, it has not been possible to properly address the possibility that CP1 expressed in *other* systems than *P. pastoris* is able to adopt a more defined-fold – in other words the lack of folding could be a *P. pastoris* –specific artefact. This seems unlikely, however, given the successful precedent for use of this expression system (see above), and its success in the production of CP2 (see below). The possible influence of the N-terminal tail of $GABA_B$ R1a on folding was also explored and it was found to have little or none. In a limited way, the importance of glycosylation was also investigated and this too was found to be minimal. It was not possible to look in detail at CP1 in the full length $GABA_B$ R1a, however, the influence on structure of the second CP (CP2) was also monitored in the course of this investigation (see below).

### *The second CP module of $GABA_B$ R1a*

Module 2 of $GABA_B$ R1a also expresses well in *P. pastoris*. In contrast with CP1, however, its HSQC and NOESY NMR spectra were very promising from the perspective of structure calculation in that they contain well-defined, discrete cross-peaks. The marked differences between modules 1 and 2 were further substantiated by other biophysical studies. For example, DSC showed a strong, sharp, transition with a relatively high melting temperature that was consistent with a cooperative unfolding event. Far-UV CD-monitored denaturation also showed cooperativity of unfolding with temperature, as expected for a protein with a stable and defined structure. Further structural studies identified the presence of two conformers for this module arising from *cis-trans* isomerisation at a peptidyl proline bond. The *cis-trans* conformers were clearly visible in the HSQC of the CP12 module pair but it is not known if isomerisation of this kind would be present in the intact receptor. Although *cis* peptidyl-Pro bonds have been noted in CP modules (e.g. factor H~5 [227]), to date *cis-trans* peptidyl-Pro isomerisation has not been reported in other CP modules. It has however been seen in isolated TB domains of fibrillins [216] and in an SH3 domain [214]. It is interesting to note that a *cis*-peptidyl-Pro bond was observed in VCP2,3 (the central modules of VCP expressed as a module pair [228]) but in the crystal structure of intact VCP, the equivalent bond was *trans*- [222]. Apart from the presence of the two conformers, CP2 looks like a regular CP module with short β-strands wrapped around a hydrophobic core, bounded at the N-terminal and C-terminal ends by the disulphides. This structural similarity is not surprising given the similarities in sequence of CP2 to classical CP-modules.

### *The CP module-pair of $GABA_B$ R1a – implications for the extent of regular folding of CP-1 in the intact receptor*

The CP12 pair from $GABA_B$ R1a was expressed in *P. pastoris*, without the N-terminal extension, by Hawrot *et al.* [34] and used as evidence for the presence of the predicted (on the basis of sequence) CP-modules on the basis of CD and DSC. Following isotopic labelling and the development of an improved purification protocol (see chapter three), this construct was shown (in the present work) to yield NMR spectra of overall poor quality. In the light of the results with the single modules, it is clear that the problem with spectra of the double module lies in the folding of module 1. The spectrum of CP12 is, interestingly, not quite the sum of the individual module spectra (i.e. poorly folded CP1 plus well-folded CP2), indicating that the modules do exert some mutual influence on one another, and the spectrum of CP1 "improves" (is suggestive of better defined structure) when it is in the context of the pair. The importance of intermodular contacts is then brought into question, however, by the observation that in the $GABA_B$ R1c subtype, CP2 is missing - that is, it appears unlikely that functional CP1 requires the presence of CP2. As with CP1, the possibility was discounted that glycosylation, or the presence of the N-terminus, might improve the spectra; moreover a range of different sample conditions also failed to improve the spectra. At this stage of the work, the choice was between attempting assignments of the module pair using triple resonance experiments on the one hand, or performing detailed studies of the more experimentally tractable CP2 on the other. It was decided to pursue the latter path with a view to returning to the double module once the second module had been fully characterised. The outcome of this work is described in chapter five.

Nevertheless the work on the double module provided an internal control in the form of CP2 – i.e. the fact that CP2 has a well-defined fold, while CP1 does not, supports the hypothesis that this property is a function of the amino acid residue sequence, and is not due to an inadequacy of the yeast protein production/folding machinery. If other domains within the $GABA_B$ R1a subunit of the $GABA_B$ heterodimer were required for correct CP folding, then intuitively it is CP2, rather than CP1, that one would expect to be perturbed by expression in the absence of the remainder of the chain. Further (circumstantial) evidence that neither CP is making very extensive, mutually stabilising contacts with the rest of the protein is that the (functional, presumably stable) $GABA_B$ R1b splice variant lacks the CP modules altogether. Therefore, taking all of these arguments together, it is reasonable to hypothesise at least that in the functional receptor, module 1 is indeed not highly structured; by extension this leads one to speculate that this low level of structure could have evolved for a specific

purpose (the sequence of CP1, unlike CP2, does indeed diverge very significantly from the sequences of the CP modules found in, for example, the RCA proteins [34]). In other words it could be critical to the biological function of the $GABA_B$ R1a subunit.

There is a growing literature precedent for the importance of intrinsically disordered proteins [229, 230]. Such proteins have been reported to exhibit molten globule-like regions that enable diverse functions or binding partners. They could represent more than 30% of eukaryotic proteins, according to a recent study on 29 genomes [230]. It has been suggested that disorder would be advantageous not only for interactions with multiple partners with varying affinities, but also to facilitate rapid response in signaling and regulation events [229]. Against this literature background, it is tempting to speculate that CP1 is indeed an intrinsically disordered domain in the context of the native $GABA_B$ receptor that would allow the receptor to interact with multiple extracellular proteins thereby providing the range of functions that this receptor has been shown to have pharmacologically.

### *The fibulin-type module*

If CP1 exhibits less conformational freedom in the presence of its binding partner(s) then this should become evident from structural studies of its complexes. To facilitate such studies, the FTM from fibulin-2 was required. Expression was attempted in *P. pastoris*, insect cells (not shown) and, extensively, in *E. coli*. Of these, *E. coli* appears to be the most promising. Further progress could be made through the development of a refolding protocol for insoluble material from inclusion material. Alternatively, the use of a fusion partner, perhaps combined with re-folding [231] might be attempted. It is difficult to explain why *P. pastoris* did not produce useful quantities of protein; amongst the possible explanations are: codon usage, uncertainties in the domain boundaries, and early transcription termination [161]. The mutual binding of intrinsically disordered proteins is a possibility for CP1 and FTM, as seen in the case of ligand-activated transcription factors [232]. These authors found that co-expression of the two components was advantageous. In their case both proteins were intracellular, but this strategy might well be worth trying for CP1 of the $GABA_B$ R1a and FTM of fibulins.

# REFERENCES


1. Smith, G.B. and R.W. Olsen, *Functional domains of $GABA_A$ receptors.* (1995) Trends Pharmacol. Sci. **16**: p. 162-168.
2. Bowery, N.G., *$GABA_B$ receptors and their significance in mammalian pharmacology.* (1989) Trends Pharmacol. Sci. **10**: p. 401-407.
3. Thompson, S.M., M. Capogna, and M. Scanziani, *$GABA_B$ receptors and their significance in mammalian pharmacology.* (1993) Trends Pharmacol. Sci. **16**: p. 222-227.
4. Kerr, D.I. and J. Ong, *$GABA_B$ receptors.* (1995) Pharmacol. Ther. **67**: p. 187-246.
5. Bettler, B., K. Kaupmann, and N. G.Bowery, *$GABA_B$ receptors: drugs meet clones.* (1998) Curr. Opin. Neurobiol. **8**: p.
6. Davies, C.H., S.J. Starkey, M.F. Pozza, and G.L. Collingridge, *$GABA_B$ autoreceptors regulate the induction of LTP.* (1991) Nature. **349**: p. 609-611.
7. Bowery, N.G., *$GABA_B$ receptor pharmacology.* (1993) Annu. Rev. Pharmacol. Toxicol. **33**: p. 109-147.
8. Caddick, S.J. and D.A. Hosford, *The role of $GABA_B$ mechanisms in animal models of absence seizures.* (1996) Mol. Neurobiol. **13**: p. 23-32.
9. Shoaib, M., L.S. Swanner, C.E. Beyer, S.R. Goldberg, and C.W. Schindler, *The $GABA_B$ agonist baclofen modifies cocaine self-administration in rats.* (1998) Behav. Pharmacol. **9**: p. 195-206.
10. Capasso, A., *$GABA_B$ receptors are involved in the control of acute opiate withdrawal in isolated tissue.* (1999) Prog. Neuropsychopharmacol. Biol. Psychiatry. **23**: p. 289-299.
11. Getova, D., N.G. Bowery, and V. Spassov, *Effects of $GABA_B$ receptor antagonists on learning and memory retention in a rat model of absence epilepsy.* (1997) Eur. J. Pharmacol. **320**: p. 9-13.
12. Hill, D.R. and N.G. Bowery, *3H-baclofen and 3H-GABA bind to bicuculline-insensitive $GABA_B$ sites in rat brain.* (1981) Nature. **290**: p. 149-152.
13. Kaupmann, K., K. Huggel, J. Heid, P.J. Flor, S. Bischoff, S.J. Mickel, G. McMaster, C. Angst, H. Bittiger, W. Froestl, and B. Bettler, *Expression cloning of GABA(B) receptors uncovers similarity to metabotropic glutamate receptors.* (1997) Nature. **386**: p. 239-246.

14. Jones, K.A., B. Borowsky, J.A. Tamm, D.A. Craig, M.M. Durkin, and M. Dai, *GABA(B) receptors function as a heteromeric assembly of the subunits GABA(B)R1 and GABA(B)R2.* (1998) Nature. **396**: p. 674-679.

15. Kaupmann, K., B. Malitschek, V. Schuler, J. Heid, W. Froestl, and P. Beck, *GABA(B)-receptor subtypes assemble into functional heteromeric complexes.* (1998) Nature. **396**: p. 683-687.

16. White, J.H., A. Wise, M.J. Main, A.Green, N.J. Fraser, and G.H. Disney, *Heterodimerization is required for the formation of a functional GABA(B) receptor.* (1998) Nature. **396**: p. 679-682.

17. Kuner, R., G. Köhr, S. Grünewald, G. Eisenhardt, A. Bach, and H.C. Kornau, *Role of heteromer formation in $GABA_B$ receptor function.* (1999) Science. **283**: p. 74-77.

18. Martin, S.C., S.J. Russek, and D.H. Farb, *Molecular identification of the human $GABA_BR2$: cell surface expression and coupling to adenylyl cyclase in the absence of $GABA_BR1$.* (1999) Mol. Cell. Neurosci. **13**: p. 180-191.

19. Ng, G.Y., T. McDonald, T. Bonnert, M. Rigby, R. Heavens, and P. Whiting, *Cloning of a novel G-protein-coupled receptor GPR 51 resembling $GABA_B$ receptors expressed predominantly in nervous tissues and mapped proximal to the hereditary sensory neuropathy type 1 locus on chromosome 9.* (1999) Genomics. **56**: p. 288-295.

20. Ng, G.Y., J. Clark, N. Coulombe, N. Ethier, T.E. Hebert, and R. Sullivan, *Identification of a $GABA_B$ receptor subunit, gb2, required for functional $GABA_B$ receptor activity.* (1999) J. Biol. Chem. **274**: p. 7607-7610.

21. Bonanno, G. and M. Raiteri, *Multiple $GABA_B$ receptors.* (1993) Trends Pharmacol. Sci. **14**: p. 259-261.

22. Kajikawa, Y., N. Saitoh, and T. Takahashi, *GTP-binding protein beta gamma subunits mediate presynaptic calcium current inhibition by $GABA_B$ receptor.* (2001) Proc. Natl. Acad. Sci. U S A. **98**: p. 8054-8058.

23. Kaupmann, K., V. Schuler, J. Mosbacher, S.Bischoff, H. Bittiger, J. Heid, W. Froestl, S. Leonhard, T. Pfaff, A. Karschin, and B. Bettler, *Human gamma-aminobutyric acid type B receptors are differentially expressed and regulate inwardly rectifying K+ channels.* (1998) Proc. Natl. Acad. Sci. U S A. **95**: p. 14991-14996.

24. Horn, F., J. Weare, M.W. Beukers, S. Horsch, A. Bairoch, and W. Chen, *GPCRDB: an information system for G protein-coupled receptors.* (1998) Nucleic Acids Res. **26**: p. 275-279.

25. Pin, J.P. and J. Bockaert, *Get receptive to metabotropic glutamate receptors.* (1995) Curr. Opin. Neurobiol. **5**: p. 342-349.

26. DeLuca, F. and J. Baron, *Molecular biology and clinical importance of the Ca(2+)-sensing receptor.* (1998) Curr. Opin. Pediatr. **10**: p. 435-440.

27. Fan, G.F., K. Ray, X.M. Zhao, P.K. Goldsmith, and A.M. Spiegel, *Mutational analysis of the cysteines in the extracellular domain of the human Ca2+ receptor: effects on cell surface expression, dimerization and signal transduction.* (1998) FEBS Lett. **436**: p. 353-356.

28. Malitschek, B., C. Schweizer, M. Keir, J. Heid, W. Froestl, and J. Mosbacher, *The N-terminal domain of gamma - aminobutyric acid B receptors is sufficient to specify agonist and antagonist binding.* (1999) Mol. Pharmacol. **56**: p. 448-454.

29. Takahashi, K., K.Tsuchida, Y. Tanabe, M. Masu, and S. Nakanishi, *Role of the large extracellular domain of metabotropic glutamate receptors in agonist selectivity determination.* (1993) J. Biol. Chem. **268**: p. 19341-19345.

30. Bräuner-Osborne, H., A.A. Jensen, P.O. Sheppard, P. O'Hara, and P. Krogsgaard-Larsen, *The agonist-binding domain of the calcium-sensing receptor is located at the amino-terminal domain.* (1999) J. Biol. Chem. **274**: p. 18382-18386.

31. Hebert, T.E., S. Moffett, J.P. Morello, T.P. Loisel, D.G. Bichet, and C. Barret, *A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation.* (1996) J. Biol. Chem. **271**: p. 16384-16392.

32. O'Hara, P.J., P.O. Sheppard, H. Thogersen, D. Venezia, B.A. Haldeman, and V.M. Grane, *The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins.* (1993) Neuron. **11**: p. 41-52.

33. Galvez, T., M.L. Parmentier, C. Joly, B. Malitschek, K. Kaupmann, and R. Kuhn, *Mutagenesis and modeling of the $GABA_B$ receptor extracellular domain support a venus flytrap mechanism for ligand binding.* (1999) J. Biol. Chem. **274**: p. 13362-13369.

34. Hawrot, E., Y. Xiao, Q.L. Shi, D. Norman, M. Kirkitadze, and P.N. Barlow, *Demonstration of a tandem pair of complement protein modules in GABA(B) receptor 1a.* (1998) FEBS Lett. **432**: p. 103-108.

35. Morley, B.J. and R.D. Campbell, *Internal homologies of the Ba fragment from human complement component Factor B, a class III MHC antigen.* (1984) EMBO J. **3**: p. 153-157.

36. Kammerer, R.A., S. Frank, T.Schulthess, R. Landwehr, A. Lustig, and J. Engel, *Heterodimerization of a functional $GABA_B$ receptor is mediated by parallel coiled-coil alpha-helices.* (1999) Biochemistry. **38**: p. 13263-13269.

37. O'Brien RJ, Lau LF, and H. RL., *Molecular mechanisms of glutamate receptor clustering at excitatory synapses.* (1998) Curr. Opin. Neurobiol.364-369.

38. Pace, A.J., L. Gama, and G.E. Breitwieser, *Dimerization of the calcium-sensing receptor occurs within the extracellular domain and is eliminated by Cys --> Ser mutations at Cys101 and Cys236.* (1999) J. Biol. Chem. **274**: p. 11629-11634.

39. Couve, A., A.K. Filippov, C.N. Connolly, B. Bettler, D.A. Brown, and S.J. Moss, *Intracellular retention of recombinant $GABA_B$ receptors.* (1998) J. Biol. Chem. **273**: p. 26361-26367.

40. Pagano, A., G. Rovelli, J. Mosbacher, T. Lohmann, B. Duthey, D. Stauffer, D. Ristig, V. Schuler, I. Meigel, C. Lampert, T. Stein, L. Prezeau, J. Blahos, J.P. Pin, W. Froestl, R. Kuhn, J. Heid, K. Kaupmann, and B. Bettler, *C-terminal interaction is essential for surface trafficking but not for heteromeric assembly of GABA(b) receptors.* (2001) J. Neurosci. **21**: p. 1189-1202.

41. Calver, A.R., R.M. J., C. Cosio, S.Q. Rice, A.J. Babbs, W.D. Hirst, I. Boyfield, M.D. Wood, R.B. Russell, G.W. Price, A. Couve, S.J. Moss, and M.N. Pangalos, *The C-terminal domains of the GABA(b) receptor subunits mediate intracellular trafficking but are not required for receptor signaling.* (2001) J. Neurosci. **21**: p. 1203-1210.

42. Margeta-Mitrovic, M., Y.N. Jan, and L.Y. Jan, *A trafficking checkpoint controls GABA(B) receptor heterodimerization.* (2000) Neuron. **27**: p. 97-106.

43. Benke, D., M. Honer, C. Michel, B. Bettler, and H. Mohler, *gamma-aminobutyric acid type B receptor splice variant proteins GBR1a and GBR1b are both associated with GBR2 in situ and display differential regional and subcellular distribution.* (1999) J. Biol. Chem. **274**: p. 27323-27330.

44. Schuler, V., C. Luscher, C. Blanchet, N. Klix, G. Sansig, K. Klebs, M. Schmutz, J. Heid, C. Gentry, L. Urban, A. Fox, W. Spooren, A. Jaton, J. Vigouret, M. Pozza, P.H. Kelly, J. Mosbacher, W. Froestl, E. Kaslin, R. Korn, S. Bischoff, K. Kaupmann, H. van der Putten, and B. Bettler, *Epilepsy, hyperalgesia, impaired memory, and loss of pre- and postsynaptic gaba(b) responses in mice lacking gaba(b(1)).* (2001) Neuron. **31**: p. 47-58.

45. Isomoto, S., M. Kaibara, Y. Sakurai-Yamashita, Y. Nagayama, Y. Uezono, and K. Yano, *Cloning and tissue distribution of novel splice variants of the rat $GABA_B$ receptor.* (1998) Biochem. Biophys. Res. Commun. **253**: p. 10-15.

46. Baldwin, J.M., *The probable arrangement of the helices in G protein-coupled receptors.* (1993) EMBO J. **12**: p. 1693-1703.

47. Bai, M., S. Trivedi, and E.M. Brown, *Dimerization of the extracellular calcium-sensing receptor (CaR) on the cell surface of CaR-transfected HEK293 cells.* (1998) J. Biol. Chem. **273**: p. 23605-23610.

48. Pfaff, T., B. Malitschek, K. Kaupmann, L. Prezeau, J.P. Pin, B. Bettler, and A. Karschin, *Alternative splicing generates a novel isoform of the rat metabotropic GABA(B)R1 receptor.* (1999) Eur. J. Neurosci. **11**: p. 2874-2882.

49. Calver, A.R., A.D. Medhurst, M.J. Robbins, K.J. Charles, M.L. Evans, D.C. Harrison, M. Stammers, S.A. Hughes, G. Hervieu, A. Couve, S.J. Moss, D.N. Middlemiss, and M.N. Pangalos, *The expression of GABA(B1) and GABA(B2) receptor subunits in the CNS differs from that in peripheral tissues.* (2000) Neuroscience. **100**: p. 155-170.

50. Schwarz, D.A., G. Barry, S.D. Eliasof, R.E. Petroski, P.J. Conlon, and R.A. Maki, *Characterization of gamma -Aminobutyric Acid Receptor $GABA_B$(1e), a $GABA_B$(1) Splice Variant Encoding a Truncated Receptor.* (2000) J. Biol. Chem. **275**: p. 32174-32181.

51. Billinton, A., N. Upton, and N.G. Bowery, *GABA(B) receptor isoforms GBR1a and GBR1b, appear to be associated with pre- and post-synaptic elements respectively in rat and human cerebellum.* (1999) Br. J. Pharmacol. **126**: p. 1387-1392.

52. Lorente, P., A. Lacampagne, Y. Pouzeratte, S. Richards, B. Malitschek, R. Kuhn, B. Bettler, and G. Vassort, *gamma -Aminobutyric acid type B receptors are expressed and functional in mammalian cardiomyocytes.* (2000) Proc. Natl. Acad. Sci. U S A. **97**: p. 8664-8669.

53. Alkadhi, K.A., D. Salgado, C.A. Davis, and E.I. Udofia, *Effects of gamma-aminobutyric acid on the compound action potential of the rat superior cervical ganglion.* . (1993) Arch. Int. Pharmacodyn. Ther. **322**: p. 66-79.

54. Feldman, D.S. and J.J. Buccafusco, *Spinal muscarinic, glutamatergic and GABAergic receptor systems in cardiovascular regulation.* (1997) J. Pharmacol. Exp. Ther. **281**: p. 274-283.

55. McKnight, A.J. and S. Gordon, *The EGF-TM7 family: unusual structures at the leukocyte surface.* (1998) J. Leukoc. Biol. **63**: p. 271-280.

56. Hamann, J., W. Eichler, D. Hamann, H.M. Kerstens, P.J. Poddighe, and J.M. Hoovers, *Expression cloning and chromosomal mapping of the leukocyte activation antigen CD97, a new seven-span transmembrane molecule of the secretion receptor superfamily with an unusual extracellular domain.* (1995) J. Immunol. **155**: p. 1942-1950.

57. Hoshino, M., E. Suzuki, T. Miyake, M. Sone, A. Komatsu, and Y. Nabeshima, *Neural expression of hikaru genki protein during embryonic and larval development of Drosophila melanogaster.* (1999) Dev. Genes Evol. **209**: p. 1-9.

58. Margolis, R.U. and R.K. Margolis, *Aggrecan-versican-neurocan family proteoglycans.* (1994) Methods Enzymol. **245**: p. 105-126.

59. Shimizu-Nishikawa, K., K. Kajiwara, and E. Sugaya, *Cloning and characterization of seizure-related gene, SEZ-6.* (1995) Biochem. Biophys. Res. Commun. **216**: p. 382-389.

60. Herbst, R. and M.J. Nicklin, *SEZ-6: promoter selectivity, genomic structure and localized expression in the brain.* (1997) Brain Res.Mol. Brain Res. **44**: p. 309-322.

61. Meindl, A., M.R. Carvalho, K. Herrmann, B. Lorenz, H. Achatz, and B. Lorenz, *A gene (SRPX) encoding a sushi-repeat-containing protein is deleted in patients with X-linked retinitis pigmentosa.* (1995) Hum. Mol. Genet. **4**: p. 2339-2346.

62. Hamann, J., C. Stortelers, E. Kiss-Toth, B. Vogel, W. Eichler, and R.A.V. Lier, *Characterization of the CD55 (DAF)-binding site on the seven-span transmembrane receptor CD97.* (1998) Eur. J. Immunol. **28**: p. 1701-1707.

63. Hamann, J., B. Vogel, G.M.v. Schijndel, and R.A.v. Lier, *The seven-span transmembrane receptor CD97 has a cellular ligand (CD55, DAF).* (1996) J. Exp. Med. **184**: p. 1185-9.

64. Sack, J.S., M.A. Saper, and F.A. Quiocho, *Periplasmic binding protein structure and function. Refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine.* (1989) J. Mol. Biol. **206**: p. 171-191.

65. Sack, J.S., S.D. Trakhanov, I.H. Tsigannik, and F.A. Quiocho, *Structure of the L-leucine-binding protein refined at 2.4 A resolution and comparison with the Leu/Ile/Val-binding protein structure.* (1989) J. Mol. Biol. **206**: p. 193-207.

66. Galvez, T., L. Prezeau, G. Milioti, M. Franek, C. Joly, W. Froestl, B. Bettler, H.-O. Bertrand, J. Blahos, and J.-P. Pin, *Mapping the Agonist-binding Site of $GABA_B$ Type 1 Subunit Sheds Light on the Activation Process of $GABA_B$ Receptors.* (2000) J. Biol. Chem. **275**: p. 41166-41174.

67. Galvez, T., S. Urwyler, L. Prezeau, J. Mosbacher, C. Joly, B. Malitschek, J. Heid, I. Brabet, W. Froestl, B. Bettler, K. Kaupmann, and J.P. Pin, *Ca(2+) requirement for high-affinity gamma-aminobutyric acid(GABA) binding at GABA(B) receptors: involvement of serine 269 of the GABA(B)R1 subunit.* (2000) Mol. Pharmacol. **57**: p. 419-426.

68. Kunishima, N., Y. Shimada, Y. Tsuji, T. Sato, M. Yamamoto, T. Kumasaka, S. Nakanishi, H. Jingami, and K. Morikawa, *Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor.* (2000) Nature. **407**: p. 971-977.

69. Livnah, O., E.A. Stura, S.A. Middleton, D.L. Johnson, L.K. Jolliffe, and I.A. Wilson, *Crystallographic Evidence for Preformed Dimers of Erythropoietin Receptor Before Ligand Activation.* (1999) Science. **283**: p. 987-990.

70. Remy, I., I.A. Wilson, and S.W. Michnick, *Erythropoietin Receptor Activation by a Ligand-Induced Conformation Change.* (1999) Science. **283**: p. 990-993.

71. Galvez, T., B. Duthey, J. Kniazeff, J. Blahos, G. Rovelli, B. Bettler, L. Prezeau, and J.-P. Pin, *Allosteric interactions between GB1 and GB2 subunits are required for optimal $GABA_B$ receptor function.* (2001) EMBO J. **20**: p. 2152-2159.

72. Malitschek, B., D. Ruegg, J. Heid, K. Kaupmann, H. Bittiger, and W. Frostl, *Developmental changes of agonist affinity at $GABA_BR1$ receptor variants in rat brain.* (1998) Mol. Cell. Neurosci. **12**: p. 56-64.

73. Marshall, F.H., K.A. Jones, K. Kaupmann, and B. Bettler, *$GABA_B$ receptors - the first 7TM heterodimers.* (1999) Trends Pharmacol. Sci. **20**: p. 396-399.

74. Mezler, M., T. Muller, and K. Raming, *Cloning and functional expression of GABA(B) receptors from Drosophila.* (2001) Eur. J. Neurosci. **13**: p. 477-486.

75. Ozenberger B.A and Y. K.H, *Investigation of ligand/receptor interactions and the formation of tertiary complexes*, in *The yeast two-hybrid system*, **bartel P.L** and **F. S**, Editors. (1997) Oxford University Press: New York. p. 158-172.

76. Mongiat, M., G. Mungiguerra, S. Bot, M.T. Mucignat, E. Giacomello, R. Doliana, and A. Colombatti, *Self-assembly and Supramolecular Organization of EMILIN.* (2000) J. Biol. Chem. **275**: p. 25471-25480.

77. Kuo, H.-J., C.L. Maslen, D.R. Keene, and R.W. Glanville, *Type VI Collagen Anchors Endothelial Basement Membranes by Interacting with Type IV Collagen.* (1997) J. Biol. Chem. **272**: p. 26522-26529.

78. Bein, K. and M. Simons, *Thrombospondin Type 1 Repeats Interact with Matrix Metalloproteinase 2. Regulation of metalloproteinase activity.* (2000) J. Biol. Chem. **275**: p. 32167-32173.

79. Blacque, O.E. and D.M. Worrall, *Evidence for a Direct Interaction between the Tumor Suppressor Serpin, Maspin, and Types I and III Collagen.* (2002) J. Biol. Chem. **277**: p. 10783-10788.

80. Mongiat, M., J. Otto, R. Oldershaw, F. Ferrer, J.D. Sato, and R.V. Iozzo, *Fibroblast Growth Factor-binding Protein Is a Novel Partner for Perlecan Protein Core.* (2001) J. Biol. Chem. **276**: p. 10263-10271.

81. Ohsawa, I., C. Takamura, and S. Kohsaka, *Fibulin-1 binds the amino-terminal head of beta-amyloid precursor protein and modulates its physiological function.* (2001) J. Neurochem. **76**: p. 1411-1420.

82. Milligan G and W. J.H., *Protein-protein interactions at G-protein-coupled receptors.* (2001) Trends Pharmacol. Sci. **22**: p. 513-518.

83. Tsui CC, Copeland NG, Gilbert DJ, Jenkins NA, Barnes C, and W. PF., *Narp, a novel member of the pentraxin family, promotes neurite outgrowth and is dynamically regulated by neuronal activity.* (1996) J. Neurosci. **16**: p. 2463-2478.

84. Fong DK and C. AM., *The Narp hypothesis?* (1999) Neuron. **23**: p. 309-323.

85. O'Brien RJ, Xu D, Petralia RS, Steward O, Huganir RL, and W. P., *Synaptic clustering of AMPA receptors by the extracellular immediate-early gene product Narp.* (1999) Neuron. **23**: p. 195-197.

86. Sasaki, T., K. Mann, H. Wiedemann, W. Gohring, A. Lustig, J. Engel, M.-L. Chu, and R. Timpl, *Dimer model for the microfibrillar protein fibulin-2 and identification of the connecting disulfide bridge.* (1997) EMBO J. **16**: p. 3035-3043.

87. Argraves, W.S., H. Tran, W.H. Burgess, and K. Dickerson, *Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure.* (1990) J. Cell Biol. **111**: p. 3155-3164.

88. Tran, H., M. Mattei, S. Godyna, and W.S. Argraves, *Human fibulin-1D: molecular cloning, expression and similarity with S1-5 protein, a new member of the fibulin gene family.* (1997) Matrix Biol. **15**: p. 479-493.

89. Adam, S., W. Gohring, H. Wiedemann, M.L. Chu, R. Timpl, and G. Kostka, *Binding of fibulin-1 to nidogen depends on its C-terminal globular domain and a specific array of calcium-binding epidermal growth factor-like (EGF) modules.* (1997) J. Mol. Biol. **272**: p. 226-236.

90. Balbona, K., H. Tran, S. Godyna, K.C. Ingham, D.K. Strickland, and W.S. Argraves, *Fibulin binds to itself and to the carboxyl-terminal heparin-binding region of fibronectin.* (1992) J. Biol. Chem. **267**: p. 20120-201205.

91. Tran, H., W.J. VanDusen, and W.S. Argraves, *The Self-association and Fibronectin-binding Sites of Fibulin-1 Map to Calcium-binding Epidermal Growth Factor-like Domains.* (1997) J. Biol. Chem. **272**: p. 22600-22606.

92. Aspberg, A., S. Adam, G. Kostka, R. Timpl, and D. Heinegard, *Fibulin-1 Is a Ligand for the C-type Lectin Domains of Aggrecan and Versican.* (1999) J. Biol. Chem. **274**: p. 20444-20449.

93. Pan, T.C., T. Sasaki, R.Z. Zhang, R. Fassler, R. Timpl, and M.L. Chu, *Structure and expression of fibulin-2, a novel extracellular matrix protein with multiple EGF-like repeats and consensus motifs for calcium binding.* (1993) J. Cell Biol. **123**: p. 1269-1277.

94. Hopf, M., W. Gohring, K. Mann, and R. Timpl, *Mapping of binding sites for nidogens, fibulin-2, fibronectin and heparin to different IG modules of perlecan.* (2001) J. Mol. Biol. **311**: p. 529-541.

95. Sasaki, T., W. Gohring, T.C. Pan, M.L. Chu, and R. Timpl, *Binding of mouse and human fibulin-2 to extracellular matrix ligands.* (1995) J. Mol. Biol. **254**: p. 892-899.

96. Olin, A.I., M. Morgelin, T. Sasaki, R. Timpl, D. Heinegard, and A. Aspberg, *The Proteoglycans Aggrecan and Versican Form Networks with Fibulin-2 through Their Lectin Domain Binding.* (2001) J. Biol. Chem. **276**: p. 1253-1261.

97. Utani, A., M. Nomizu, and Y. Yamada, *Fibulin-2 Binds to the Short Arms of Laminin-5 and Laminin-1 via Conserved Amino Acid Sequences.* (1997) J. Biol. Chem. **272**: p. 2814-2820.

98. Giltay, R., R. Timpl, and G. Kostka, *Sequence, recombinant expression and tissue localization of two novel extracellular matrix proteins, fibulin-3 and fibulin-4.* (1999) Matrix Biol. **18**: p. 469-480.

99. Gallagher, W.M., L.M. Greene, M.P. Ryan, V. Sierra, A. Berger, P. Laurent-Puig, and E. Conseiller, *Human fibulin-4: analysis of its biosynthetic processing and mRNA expression in normal and tumour tissues.* (2001) FEBS Lett. **489**: p. 59-66.

100. Kowal, R.C., J.A. Richardson, J.M. Miano, and E.N. Olson, *EVEC, a Novel Epidermal Growth Factor–Like Repeat-Containing Protein Upregulated in Embryonic and Diseased Adult Vasculature.* (1999) Circ. Res. **84**: p. 1166-1176.

101. Nakamura, T., P. Ruiz-Lozano, V. Lindner, D. Yabe, M. Taniwaki, Y. Furukawa, K. Kobuke, K. Tashiro, Z. Lu, N.L. Andon, R. Schaub, A. Matsumori, S. Sasayama, K.R. Chien, and T. Honjo, *DANCE, a Novel Secreted RGD Protein Expressed in Developing, Atherosclerotic, and Balloon-injured Arteries.* (1999) J. Biol. Chem. **274**: p. 22476-22483.

102. Yanagisawa H, Davis EC, Starcher BC, Ouchi T, Yanagisawa M, Richardson JA, and O. EN., *Fibulin-5 is an elastin-binding protein essential for elastic fibre development in vivo.* (2002) Nature. **415**: p. 168-171.

103. Nakamura T, Lozano PR, Ikeda Y, Iwanaga Y, Hinek A, Minamisawa S, Cheng CF, Kobuke K, Dalton N, Takada Y, Tashiro K, Ross Jr J, Honjo T, and C. KR., *Fibulin-5/DANCE is essential for elastogenesis in vivo.* (2002) Nature. **415**: p. 171-175.

104. Sasaki, T., G. Kostka, W. Gohring, H. Wiedemann, K. Mann, M.L. Chu, and R. Timpl, *Structural characterization of two variants of fibulin-1 that differ in nidogen affinity.* (1995) J. Mol. Biol. **245**: p. 241-250.

105. Perbal, B., C. Martinerie, R. Sainson, M. Werner, B. He, and B. Roizman, *The C-terminal domain of the regulatory protein NOVH is sufficient to promote interaction with fibulin 1C: A clue for a role of NOVH in cell-adhesion signaling.* (1999) Proc. Natl. Acad. Sci. U S A. **96**: p. 869-874.

106. Reinhardt, D.P., T. Sasaki, B.J. Dzamba, D.R. Keene, M.-L. Chu, W. Gohring, R. Timpl, and L.Y. Sakai, *Fibrillin-1 and Fibulin-2 Interact and Are Colocalized in Some Tissues.* (1996) J. Biol. Chem. **271**: p. 19489-19496.

107. Handford, P.A., A.K. Downing, D.P. Reinhardt, and L.Y. Sakai, *Fibrillin: from domain structure to supramolecular assembly.* (2000) Matrix Biol. **19**: p. 457-470.

108. Reinhardt, D.P., J.E. Gambee, R.N. Ono, H.P. Bachinger, and L.Y. Sakai, *Initial Steps in Assembly of Microfibrils. Formation of disulfide-cross-linked multimers containing fibrillin-1.* (2000) J. Biol. Chem. **275**: p. 2205-2210.

109. Milewicz, D.M., J. Grossfield, S.N. Cao, C. Kielty, W. Covitz, and T. Jewett, *A mutation in FBN1 disrupts profibrillin processing and results in isolated skeletal features of the Marfan syndrome.* (1995) J. Clin. Invest. . **95**: p. 2373-2378.

110. Ritty, T.M., T. Broekelmann, C. Tisdale, D.M. Milewicz, and R.P. Mecham, *Processing of the Fibrillin-1 Carboxyl-terminal Domain.* (1999) J. Biol. Chem. **274**: p. 8933-8940.

111. Neuhaus D. and W. M., *The Nuclear Overhauser Effect in structural and conformational analysis.* (1989) VCH.

112. Freeman, R., *A Handbook of Nuclear Magnetic Resonance - 2nd Edition.* (1997) LONGMAN Publishers Ltd.

113. Ven, F.J.M.v.d., *Multidimensional NMR in Liquids : Basic Principles and experimental methods.* (1995) WILEY-VCH.

114. Verdun F. R. and A.G. Marshall, *Tranformee de Fourier: Applications en RMN et IRM.* (1995) Masson-Paris-France.

115. Lindon J.C. and F. A.G., *Digitisation and data processing in Fourier transform NMR.* (1980) Prog. NMR Spec. **14**: p. 27-66.

116. Delsuc M.A., Fortier P.L.., van Heijenoort C., Malliavin T., Robin M., and R. A., *Le traitement des donnees en RMN.* (1995) Cahiers Ecole IMABIO RMN des Proteines -CNRS-France49-61.

117. Delsuc M.A. and L. G.C., *The application of Maximum Entropy processing to the deconvolution of coupling patterns in NMR.* (1988) J. Magn. Reson. **76**: p. 306-315.

118. Laue E.D., Skilling J, Staunton J., Sibisi S., and B. R.G., *Maximum Entropy Method in Nuclear Magnetic Resonance Spectroscopy.* (1985) J. Magn. Reson. **62**: p. 437-452.

119. Mo H. and P. T.C., *Intermolecular interactions characterized by nuclear Overhauser effects.* (1997) Prog. NMR Spec. **30**: p. 1-38.

120. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, and S. BD., *1H, 13C and 15N chemical shift referencing in biomolecular NMR.* (1995) J.Biomol. NMR. **6**: p. 135-140.

121. Ernst R.R., Bodenhausen G., and W. A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* International Series of Monographs on Chemistry 14. (1987) Oxford science publication.

122. Wimperis S. and H. P.J., *Eight lectures on NMR Spectroscopy.* (1992) Cambrigde Unversity - UK.

123. Waugh, J.S., *Theory of Broadband Spin Decoupling.* (1982) J. Magn. Reson. **50**: p. 30-49.

124. Levitt M.H. and R. Freeman, *Composite pulse decoupling.* (1981) J. Magn. Reson. **43**: p. 502-507.

125. Brown SC, Weber PL, and M. L., *Toward complete 1H NMR spectra in proteins.* (1988) J. Magn. Reson. **77**: p. 166-169.

126. Plateau P. and G. M., *Exchangeable proton nmr without base-line distortion, using new strong-pulse sequences.* (1982) J. Am.Chem.Soc. **104**: p. 7310-7311.

127. Piotto M., Saudeck V., and S. V, *Gradient-tailored excitation for single-quantum NMR spectroscopy in aqueous solutions.* (1992) J.Biomol. NMR. **2**: p. 661-665.

128. Hwang T.L. and S. A.J., *Water suppression that works . Excitation sculpting using arbitrary waveforms and pulsed field gradients.* (1995) J. Magn. Reson. A. **112**: p. 275-279.

129. Bodenhausen G., Kogler H., and E. R.R., *Selection of Coherence-Transfer Pathways in NMR Pulse Experiments.* (1984) J. Magn. Reson. **58**: p. 370-388.

130. Aue W.P., Bartholdi E., and E. R.R., *Two-dimensional spectroscopy. Application to nuclear magnetic resonance.* (1976) J. Chem. Phys. **64**: p. 2229-2246.

131. Perczel, A., *Selected chapters of modern NMR-spectroscopy.* (1995-1997) OCMS - University of Oxford.

132. Braunschweiler L. and E. R.R., *Coherence transfert by isotropic mixing: application to protn correlation spectroscopy.* (1983) J. Magn. Reson. **53**: p.

133. Davies D.G. and B. A., *Assignment of complex 1H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy.* (1985) J. Am.Chem.Soc. **107**: p. 2820-2821.

134. Cavanagh J and R. M., *Suppression of cross relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence.* (1992) J. Magn. Reson. **96**: p. 670-678.

135. Jeener J., Meier B.H., Bachmann P., and E. R.R., *Investigation of exchange processes by two-dimensional NMR spectroscopy.* (1979) J. Chem. Phys. **71**: p. 4546-4553.

136. Bodenhausen G. and R. D.J., *Natural abudance nitrogen-15 NMR by enhanced heteronuclear spectroscopy.* (1988) Chem. Phys. Lett. **69**: p. 185-189.

137. Muller, L., *Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence.* (1979) J. Am.Chem.Soc. **101**: p. 4481-4484.

138. Palmer III, A.G., Cavanagh J., Wright P.E., and R. M., *Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation spectroscopy.* (1991) J. Magn. Reson. **93**: p. 151-170.

139. Marion D, Kay LE, Sparks SW, Torchia DA, and B. A., *Three-dimensional heteronuclear NMR of 15N-labeled protein.* (1989) J. Am.Chem.Soc.1515- 1517.

140. Mori, S., C. Abeygunawardana, M.O. Johnson, and P.C.M. van Zijl, *Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection that avoids water saturation.* (1995) J. Magn. Reson. B. **108**: p. 94-98.

141. Sklenar V, Piotto M., and L. R., *Gradient-tailored water suppression for H-1-N-15 hsqc experiments optimized to retain full sensitivity.* (1993) J. Magn. Reson. A. **102**: p. 241-245.

142. Uhrin, D. NMR lectures. (1998), Edinburgh University - UK.

143. Sattler M., Schleucher J., and G. C., *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients.* (1999) Prog. NMR Spec. **34**: p. 93-158.

144. Ducruix, A. and G. R, *Crystallisation of Nucleic Acids and Proteins - Second Edition.* The Practical Approach Series, ed. B.D. Hames. (1999) OXFORD University Press.

145. Edwards, A.M., C.H. Arrowsmith, D. Christendat, A. Dharamsi, J.D. Friesen, J.F. Greenblatt, and M. Vedadi, *Protein production: feeding the crystallographers and NMR spectroscopists.* (2000) Nat. Struct. Biol. **7**: p. 970-972.

146. Christendat, D., A. Yee, A. Dharamsi, Y. Klugerc, M. Gersteinc, C.H. Arrowsmith, and A.M. Edwardsa, *Structural proteomics: prospects for high throughput sample preparation.* (2000) Prog. Biophys. Mol. Biol. **73**: p. 339-345.

147. Montelione GT, Zheng D, Huang YJ, Gunsalus KC, and S. T., *Protein NMR spectroscopy in structural genomics.* (2000) Nat. Struct. Biol. **7**: p. 982-985.

148. Clark, E.D.B., *Refolding of recombinant proteins.* (1998) Curr. Opin. Biotechnol. **9**: p. 157-163.

149. Lilie, H., E. Schwarz, and R. Rudolph, *Advances in refolding of proteins produced in E. coli.* (1998) Curr. Opin. Biotechnol. **9**: p. 497-501.

150. *pET System Manual.* (TB055 9th Edition 05/00) Novagen.

151. Williams J.A., Langeland J.A., Thalley B.S., Skeath J.B., and C. S.B., *Expression of foreign proteins in E. coli using plasmid vector and purification of specific polyclonal antibodies*, in *DNA Cloning 2 : Expression systems - Second Edition*, D.M.Glover and B.D. Hames, Editor. (1995) IRL Press - Oxford Unversity Press. p. 15-58.

152. Georgiou, G. and P. Valax, *Isolating inclusion bodies from bacteria.* (1999) Methods Enzymol. **309**: p. 48-58.

153. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* (1988) Genes. **67**: p. 31-40.

154. LaVallie E.R, DiBlasio E.A, Kovacic S, Grant K.L, Schendel P.F, and M. J.M., *A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm.* (1993) Bio/Technology. **11**: p. 187-193.

155. Davis G.D., Elisee C, Newham D.M, and H. R.G., *New fusion protein systems designed to give soluble expression in Escherichia coli.* (1999) Biotechnol. Bioeng. **65**: p. 382-388.

156. Rietsch, A., D. Belin, N. Martin, and J. Beckwith, *An in vivo pathway for disulfide bond isomerization in Escherichia coli.* (1996) Proc. Natl. Acad. Sci. U S A. **93**: p. 13048-13053.

157. Patel G. and J. N.C., *The baculovirus expression system*, in *DNA Cloning 2 : Expression systems - Second Edition*, D.M.Glover and B.D. Hames, Editor. (1995) IRL PRESS - Oxford Unversity Press. p. 205-244.

158. Romanos M. A., Scorer C. A., and C.J. J., *Expression of cloned genes in yeast*, in *DNA Cloning 2 : Expression systems - Second Edition*, D.M.Glover and B.D. Hames, Editor. (1995) IRL Press - Oxford Unversity Press. p. 123-167.

159. White CE, Kempi NM, and K. EA., *Expression of highly disulfide-bonded proteins in Pichia pastoris.* (1994) Structure. **2**: p. 1003-1005.

160. Cereghino, J.L. and J.M. Cregg, *Heterologous protein expression in the methylotrophic yeast Pichia pastoris.* (2000) FEMS Microbiology Reviews. **24**: p. 45-66.

161. *Pichia Expression Kit - A Manual of Methods for Expression of Recombinant Proteins in Pichia pastoris*. Version J #25-0043, ed. Invitrogen.

162. Sambrook, J., E.F. Fritsch, and T. Maniatis, *Molecular cloning: a laboratory manual.*, ed. C.S.H. Laboratory. (1989) Cold spring Harbor, New York.

163. *pPICZ α A, B, and C manual - Version D -Catalog no. V195-20* Invitrogen.

164. Kay L.E., Nicholson L. K., Delagio F., Bax A., and T. D.A., *Pulse sequences for removal of the effects of cross relaxation between dipolar and chemical-shift anisotropy - Relaxation mechanisms on the measurement of heteronuclear $T_1$ and $T_2$ values in proteins.* (1992) J. Magn. Reson. **97**: p. 359-375.

165. Wormald MR and D. RA., *Glycoproteins: glycan presentation and protein-fold stability.* (1999) Structure. **7**: p. 155-160.

166. Wyss DF and W. G., *The structural role of sugars in glycoproteins.* (1996) Curr. Opin. Biotechnol. **7**: p. 409-416.

167. Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, Arulanandam AR, Smolyar A, Reinherz EL, and W. G., *Conformation and function of the N-linked glycan in the adhesion domain of human CD2.* (1995) Science. **269**: p. 1273-1278.

168. Pfuhl M, Improta S, Politou A. S., and P. A., *When a Module is Also a Domain: The Rôle of the N Terminus in the Stability and the Dynamics of Immunoglobulin Domains from Titin.* (1997) J. Mol. Biol. **265**: p. 242-256.

169. Politou AS, Gautel M, Joseph C, and P. A., *Immunoglobulin-type domains of titin are stabilized by amino-terminal extension.* (1994) FEBS Lett. **352**: p. 27-31.

170. Wishart D.S., Sykes B.D., and R. F.M., *Relationship between nuclear magnetic resonance chemical shift and protein secondary structure.* (1991) J. Mol. Biol. **222**: p. 311-333.

171. Schulman B.A., Kim P.S., Dobson C.M., and R. C., *A residue-specific NMR view of the non-cooperative unfolding of a molten globule.* (1997) Nat. Struct. Biol. **4**: p. 630-634.

172. Vassilenko KS and U. VN., *Native-like secondary structure of molten globules.* (2002) B.B.A. **1594**: p. 168-177.

173. Dyson HJ and W. PE., *Equilibrium NMR studies of unfolded and partially folded proteins.* (1998) Nat. Struct. Biol. **5 Suppl:**: p. 499-503.

174. Redfield, C., *Molten globules.* (1999) Curr. Biol. **9**: p. R313.

175. Jerala, R., E. Zerovnik, J. Kidric, and V. Turk, *pH-induced Conformational Transitions of the Propeptide of Human Cathepsin L. A role for a molten globule state in zymogen activation.* (1998) J. Biol. Chem. **273**: p. 11498-11504.

176. Kelly SM and P. NC., *The application of circular dichroism to studies of protein folding and unfolding.* (1997) B.B.A. **1338**: p. 161-185.

177. Greenfield, N.J., *Methods to estimate the conformation of proteins and polypeptides from circular dichroism data.* (1996) Anal. Biochem. **235**: p. 1-10.

178. Greenfield, N.J., *Applications of circular dichroism in protein and peptide analysis.* (1999) Trends Anal. Chem. **18**: p. 236-244.

179. Kirkitadze M.D., Krych M., Uhrin D., Dryden D.T., Smith B.O., Cooper A., Wang X., Hauhart R., Atkinson J.P., and B. P.N., *Independently melting modules and highly structured intermodular junctions within complement receptor type 1.* (1999) Biochemistry. **38**: p. 7019-7031.

180. Kirkitadze M.D., Henderson C., Price N.C., Kelly S.M., Mullin N.P., Parkinson J., Dryden D.T., and B. P.N., *Central modules of the vaccinia virus complement control protein are not in extensive contact.* (1999) Biochem. J. **344**: p. 167-175.

181. Plum GE and B. KJ., *Calorimetry of proteins and nucleic acids.* (1995) Curr. Opin. Struct. Biol. **5**: p. 682-690.

182. Cooper, A., *Thermodynamic analysis of biomolecular interactions.* (1999) Curr. Opin. Struct. Biol. **3**: p. 557-563.

183. Cooper A., Nutley M. A., and W. A., *Differential scanning microcalorimetry*, in *Protein-Ligand Interactions: hydrodynamics and calorimetry.* (2000) Oxford University Press, Oxford New York. p. 287-318.

184. Kirkitadze, M.D., P.N. Barlow, N.C. Price, S.M. Kelly, C.J. Boutell, F.J. Rixon, and D.A. McClelland, *The Herpes Simplex Virus Triplex Protein, VP23, Exists as a Molten Globule.* (1998) J. Virol. **72**: p. 10066-10072.

185. Ladokhin, A.S., *Fluorescence spectroscopy in peptide and protein analysis*, in *Encyclopedia of Analytical Chemistry*, R.A. Meyers, Editor. (2000) John Wiley & Sons Ltd. p. 5762-5779.

186. Kirkitadze M.D., Jumel K., Harding S.E., Dryden D.T., Krych M., Atkinson J.P., and P.N. Barlow, *Combining ultracentrifugation with fluorescence to follow the unfolding of modules 16-17 of the complement receptor type 1.* (1999) Prog. Colloid Polymer Sci. **113**: p. 164-167.

187. Bailey R.W., Dunker A.K., Brown C.J., Garner E.C., and G. M.D., *Clusterin, a binding protein with a molten globule-like region.* (2001) Biochemistry. **40**: p. 11828-11840.

188. Harding S.E. and W. D.J., *Sedimentation velocity analytical ultracentrifugation*, in *Protein-Ligand Interactions: hydrodynamics and calorimetry*, S. E. Harding and B. Z. Chowdhry, Editors. (2000) Oxford University Press, Oxford New York. p. 75-135.

189. Laue, T., *Biophysical studies by ultracentrifugation.* (2001) Curr. Opin. Struct. Biol.579-583.

190. Santos, N.C. and M.A. Castanho, *Teaching Light Scattering Spectroscopy: The Dimension and Shape of Tobacco Mosaic Virus.* (1996) Biophys. J. **71**: p. 1641-1650.

191. Kraulis P.J., Domaille P.J., Campbell-Burk S.L., Van Aken T., and L. E.D., *Solution structure and dynamics of ras p21.GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy.* (1994) Biochemistry. **33**: p. 3515-3531.

192. Rance, M., *Improved techniques for homonuclear rotating-frame and isotropic mixing experiments.* (1987) J. Magn. Reson. **74**: p. 557-564.

193. Pascal, S.M., D.R. Muhandiram, T. Yamazaki, J.D. Forman-Kay, and L.E. Kay, *Simultaneous acquisition of N-15-edited and C-13-edited NOE spectra of proteins dissolved in H2O.* (1994) J. Magn. Reson. B. **103**: p. 197-201.

194. Aitio H and P. P., *Semi-constant-time HMSQC (SCT-HMSQC-HA) for the measurement of (3)J(H(N))(H(alpha)) couplings in (15)N-labeled proteins.* (2000) J. Magn. Reson. **143**: p. 391-396.

195. Dux P., Whitebread B., Boelens R., Kaptein R., and V.G. W., *Measurement of 15 N-1H coupling constants in uniformly 15N-labeled proteins: Application to the photoactive yellow protein.* (1997) J.Biomol. NMR. **10**: p. 301-306.

196. Mori, S., C. Abeygunawardana, J.M. Berg, and P.C.M. van Zijl, *NMR study of rapidly exchanging backbone amide protons in staphylococcal nuclease and the correlation with structural and dynamic properties.* (1997) J. Am.Chem.Soc. **119**: p. 6844-6852.

197. Vuister G.W. and B. A, *Resolution enhancement and spectral editing of uniformly C-13-enriched proteins by homonuclear broad-band C-13 decoupling.* (1992) J. Magn. Reson. **98**: p. 428-435.

198. Kay, L.E., G.Y. Xu, and T. Yamazaki, *Enhanced-sensitivity triple-resonance spectroscopy with minimal H2O saturation.* (1994) J. Magn. Reson. A. **109**: p. 129-133.

199. Engelke, J. and H.J. Ruterjans, *Sequential protein backbone resonance assignments using an improved 3D-HN(CA)CO pulse scheme.* (1995) J. Magn. Reson. B. **109**: p. 318-322.

200. Muhandiram, D.R. and L.E. Kay, *Gradient enhanced triple resonance three-dimensional NMR experiments with improved sensitivity.* (1994) J. Magn. Reson. B. **103**: p. 203-216.

201. Grzesiek S. and B. A, *Amino-acid type determination in the sequential assignment procedure of uniformly C-13/N-15-enriched proteins.* (1993) J.Biomol. NMR. **3**: p. 185-204.

202. Wang A.C., Lodi P.J., and Q. J., *An efficient triple-resonance experiment for proton-directed sequential backbone assignment of medium-sized proteins.* (1994) J. Magn. Reson. B. **105**: p. 196-198.

203. Grzesiek S., Anglister J., and B. A., *Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization.* (1993) J. Magn. Reson. B. **101**: p. 114-119.

204. Kay, L.E., G.Y. Xu, A.U. Singer, D.R. Muhandiram, and J.D. Forman-Kay, *A gradient-enhanced HCCH-TOCSY experiment for recording side-chain H-1 and C-13 correlations in H2O samples of proteins.* (1993) J. Magn. Reson. B. **101**: p. 333-337.

205. Uhrin D, Uhrinova S, Leadbeater C, Nairn J, Price NC, and B. PN., *3D HCCH(3)-TOCSY for resonance assignment of methyl-containing side chains in (13)C-labeled proteins.* (2000) J. Magn. Reson. **142**: p. 288-293.

206. Yamazaki T, Forman-Kay JD, and K. LE., *Two-dimensional NMR experiments for correlating $^{13}C$ beta and $^{1}H$ delta/epsilon chemical shifts of aromatic residues in $^{13}C$-labeled proteins via scalar couplings.* (1993) J. Am.Chem.Soc. **115**: p. 11054- 11055.

207. Evans, J.N.S., *Biomolecular NMR Spectroscopy.* (1995) Oxford University Press.

208. Wakefield RI, Smith BO, Nan X, Free A, Soteriou A, Uhrin D, Bird AP, and B. PN., *The solution structure of the domain from MeCP2 that binds to methylated DNA.* (1999) J. Mol. Biol. **291**: p. 1055-1065.

209. Nilges M, Macias MJ, O'Donoghue SI, and O. H., *Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin.* (1997) J. Mol. Biol. **269**: p. 408-422.

210. Smith BO, Mallin RL, Krych-Goldberg M, Wang X, Hauhart RE, Bromek K, Uhrin D, Atkinson JP, and B. PN., *Structure of the C3b Binding Site of CR1 (CD35), the Immune Adherence Receptor.* (2002) Cell. **108**: p. 769-780.

211. Nilges, M., *Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities.* (1995) J. Mol. Biol. **245**: p. 645-660.

212. Brunger, A.T., *X-PLOR Version 3.1.* A system for X-ray crystallography and NMR. (1992) Yale University Press.

213. Kordel J, Forsen S, Drakenberg T, and C. WJ, *The rate and structural consequences of proline cis-trans isomerization in calbindin D9k: NMR studies of the minor (cis-Pro43) isoform and the Pro43Gly mutant.* (1990) Biochemistry. **29**: p. 4400-4409.

214. Ogura K, Nagata K, Horiuchi M, Ebisui E, Hasuda T, Yuzawa S, Nishida M, Hatanaka H, and I. F., *Solution structure of N-terminal SH3 domain of Vav and the recognition site for Grb2 C-terminal SH3 domain.* (2002) J.Biomol. NMR. **22**: p. 37-46.

215. Amodeo P, Morelli MA, and C.M. A., *Multiple conformations and proline cis-trans isomerization in salmon calcitonin: a combined nuclear magnetic resonance, distance geometry, and molecular mechanics study.* (1994) Biochemistry. **33**: p. 10754-10762.

216. Yuan X, Werner JM, Knott V, Handford PA, Campbell ID, and D. K., *Effects of proline cis-trans isomerization on TB domain secondary structure.* (1998) Protein Sci. **7**: p. 2127-2135.

217. Wishart DS and S. B.D., *The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data.* (1994) J.Biomol. NMR. **4**: p. 171-180.

218. Kabsch W. and S. C., *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* (1983) Biopolymers. **22**: p. 2577-2637.

219. Kraulis, P.J., *MOLSCRIPT: a program to produce both de-tailed and schematic plots of protein structures.* (1991) J. Appl. Crystallogr. **24**: p. 9446-9950.

220. Brasher SV, Smith BO, Fogh RH, Nietlispach D, Thiru A, Nielsen PR, Broadhurst RW, Ball LJ, Murzina NV, and L. ED., *The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer.* (2000) EMBO J. **19**: p. 1587-1597.

221. Wei K, Jia Z, Wang YT, Yang J, Liu CC, and S.O. 3rd., *Cloning and characterization of a novel variant of rat GABA(B)R1 with a truncated C-terminus.* (2001) Brain Res.Mol. Brain Res. **89 (1-2)**: p. 103-110.

222. Murthy KH, Smith SA, Ganesh VK, Judge KW, Mullin N, Barlow PN, Ogata CM, and K. GJ., *Crystal structure of a complement control protein that regulates both pathways of complement activation and binds heparan sulfate proteoglycans.* (2001) Cell. **104**: p. 301-311.

223. Lin F, I.R., Shoham M, Medof ME., *Bulk production and functional analyses of mouse CD55's native and deglycosylated active domains.* (2001) Arch. Biochem. Biophys. **393**: p. 67–72.

224. Morgan WD, Birdsall B, Frenkiel TA, Gradwell MG, Burghaus PA, Syed SE, Uthaipibull C, Holder AA, and F. J., *Solution structure of an EGF module pair from the Plasmodium falciparum merozoite surface protein 1.* (1999) J. Mol. Biol. **289**: p. 113-122.

225. Thielens, N.M., K. Enrie, M. Lacroix, M. Jaquinod, J.-F. Hernandez, A.F. Esser, and G.J. Arlaud, *The N-terminal CUB-Epidermal Growth Factor Module Pair of Human Complement Protease C1r Binds Ca2+ with High Affinity and Mediates Ca2+-dependent Interaction with C1s.* (1999) J. Biol. Chem. **274**: p. 9149-9159.

226. Clark NS, Dodd I, Mossakowska DE, Smith RA, and G. MG., *Folding and conformational studies on SCR1-3 domains of human complement receptor 1.* (1996) Protein Eng. **9**: p. 877-884.

227. Barlow PN, Norman DG, Steinkasserer A, Horne TJ, Pearce J, Driscoll PC, Sim RB, and C. ID., *Solution structure of the fifth repeat of factor H: a second example of the complement control protein module.* (1992) Biochemistry. **31**: p. 3626-3634.

228. Henderson CE, Bromek K, Mullin NP, Smith BO, Uhrin D, and B. PN, *Solution structure and dynamics of the central CCP module pair of a poxvirus complement control protein.* (2001) J. Mol. Biol. **307**: p. 323-339.

229. Dyson HJ and W. PE., *Coupling of folding and binding for unstructured proteins.* (2002) Curr. Opin. Struct. Biol. **12**: p. 54-60.

230. Dunker A.K., Lawson J.D., Brown C.J., Williams R.M., Romero P, Oh J.S., Oldfield C.J., Campen A.M., Ratliff C.M., Hipps K.W., Ausio J, Nissen M.S., Reeves R, Kang C, Kissinger C.R., Bailey R.W., Griswold M.D., Chiu W, Garner E.C., and O. Z., *Intrinsically disordered protein.* (2001) J. Mol. Graph. Model. **19**: p. 26-59.

231. Perdikoulis MV, Kishore U, and R. KB., *Expression and characterisation of the thrombospondin type I repeats of human properdin.* (2001) B.B.A. **1548**: p. 265-277.

232. Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM, and W. PE., *Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators.* (2002) Nature. **415**: p. 549-553.

# APPENDICES

## APPENDIX A: *Pichia pastoris* media recipes & protocols

### A-1 stock solutions

10 x YNB: 13.4% yeast nitrogen base with ammonium sulfate without amino acids.

$^{15}$N-10 x YNB: 13.4% yeast nitrogen base without ammonium sulfate and amino acids, and 0.2% $^{15}$N-ammonium sulfate.

500 x B: 0.02% biotin.

100 x H: 0.4% L-histidine.

10 x D: 20% dextrose.

10 x M: 5% methanol.

10 x GY: 10% glycerol.

(all solutions sterile by autoclaving or filter sterilisation)

### A-2 solutions for transformation and maintenance

YPD:yeast peptone dextrose medium

1% yeast extract
2% peptone
2% dextrose
20 g of agar per litre if making plates
Add 1 x D after autoclaving

MD and MDH: minimal dextrose medium +/- histidine

1.34% YNB
4 x $10^{-5}$% biotin
2% dextrose
Autoclave water then add 1 x YNB, 1 x D, 1 x B at 60°C
To make MDH, add 1 x H
For plates, add 15 g agar per litre

YPDS + Zeocin Agar:yeast extract peptone dextrose medium with sorbitol

1% yeast extract
2% peptone
2% dextrose
1 M sorbitol
+ 2% agar
+ appropriate concentration of Zeocin ( added at 60°C)
Autoclave water then add 1 x D

## A-3 growth solutions

100 mM potassium phosphate, pH 6.0
1.34% YNB
4 x $10^{-5}$% biotin
1% glycerol (for BMG: buffered minimal glycerol medium) or 0.5% methanol (for BMM: buffered minimal methanol medium).

Autoclave water and cool to room temperature, then add potassium phosphate buffer, pH 6.0, 1 x YNB,1 x B,1 x GY or M

For $^{15}N$ labelled growth use $^{15}N$-10 x YNB

For $^{15}N/^{13}C$ labelled growth use:

1-BMD: buffered minimal dextrose medium

100 mM potassium phosphate, pH6.0
1.34% $^{15}N$-YNB
4 x $10^{-5}$% biotin
0.4-0.5% $^{13}C$-glucose

2-BMG: buffered minimal glycerol medium (low concentration)

100 mM potassium phosphate, pH 6.0
1.34% $^{15}N$-YNB
4 x $10^{-5}$% biotin
0.1-0.05% $^{13}C$-glycerol

3-BMM: Buffered Minimal Methanol Medium

100 mM potassium phosphate, pH 6.0
1.34% $^{15}N$-YNB
4 x $10^{-5}$% biotin
0.5% $^{13}C$-methanol

## A-4 induction tests

For expression tests, individual colonies from transformation plates were used to inoculate 5-10 ml of BMG and grown for 48 h at 30°C-250 RPM. Cells were harvested (1500 g-5 min-room temperature) and resuspended in 1-2 ml of BMM for induction. Induction period was two days with 0.5% methanol final concentration maintained throughout by daily addition of 100% methanol stock.

After induction, 1.5 ml of cells was transferred into an Eppendorf tube and pelleted at 13,000 RPM for 10 min at room temperature. Supernatant was decanted in clean eppendorf tube, and both cell pellet and supernatant were assayed by SDS-PAGE. Protein constructs bearing consensus N-glycosylation site(s) were subjected to $EndoH_f$ treatment (2500 units per ml of supernatant) prior SDS-PAGE.

## APPENDIX B: oligonucleotides sequences

Note: all sequences are 5' to 3, and annealing temperature was 55°C in all PCR amplifications.

### B-1 yeast constructs

#### B-1.1 TR-CP12

[forward *EcoRI*]

CCGGAATTCGCCACCTCGGAAGGTTGCCA

[reverse N*otI*]

ATAGTTTAGCGGCCGCTCATTAATTCACCTGGCAGTGGGGCT

#### B-1.2 FLM-CP12

[forward *EcoRI*]

CCGGAATTCGGTGGTGCTCAAACTCCAGCTGCTACTTCGGAAGGTTGCCAGATTAT

ACATCCGCCC

[reverse N*otI*]

same as TR-CP12

#### B-1.3 WT-CP12

forward *EcoRI*]

CCGGAATTCGGTGGTGCTCAAACTCCAAATG

[reverse N*otI*]

same as TR-CP12

### B-1.4 TR-CP1 & WT-CP1

[forward *EcoRI*]

same as TR-CP12 & WT-CP12, respectively.

[reverse N*otI*]

ATAGTTTAGCGGCCGCTCATTAGATTCGGACACAGCGGC

### B-1.5 CP2

[forward *EcoRI*]

CCGGAATTCGTCCGAATCTGCTCCAA

[reverse N*otI*]

same as TR-CP12

## B-2 *E. coli* constructs

[forward *NdeI*]

GGAATTCCATATGGAGCGCACCACGTGCC

[reverse *XhoI*]

CCGCTCGAGTCATTACAGGGCAAAGGTGG

**APPENDIX C: assignment table of the two conformers of the $GABA_B$ R1a CP2**

| "Trans" | | "Cis" |
|---|---|---|
| 92 | Glu | |
| N | - | |
| HN | - | |
| CA | 55.7 | |
| HA | 4.04 | |
| CO | 172.3 | |
| CB | 30.1 | |
| HB1 | 2.13 | |
| HB2 | 2.08 | |
| CG | 35.3 | |
| HG1 | - | |
| HG2 | - | |
| CD | - | |
| HB# | 2.11 | |
| HG# | 2.41 | |
| 93 | Ala | |
| N | 126.2 | |
| HN | 8.66 | |
| CA | 52.6 | |
| HA | 4.31 | |
| CO | 177 | |
| CB | 19.3 | |
| HB# | 1.34 | |
| 94 | Glu | |
| N | 120.8 | |
| HN | 8.35 | |
| CA | 56.5 | |
| HA | 4.23 | |
| CO | 175.9 | |
| CB | 30.1 | |
| HB1 | 1.92 | |
| HB2 | 1.87 | |
| CG | 35.4 | |
| HG1 | 2.25 | |
| HG2 | 2.2 | |
| CD | - | |
| HB# | 1.89 | |
| HG# | 2.22 | |
| 95 | Phe | |
| N | 121.4 | |
| HN | 8.15 | |
| CA | 57.4 | |
| HA | 4.65 | |
| CO | 175.2 | |
| CB | 39.7 | |
| HB1 | 3.1 | |
| HB2 | 3.01 | |
| CD1 | - | |
| HD1 | - | |
| CD2 | - | |
| HD2 | - | |
| CE1 | - | |
| HE1 | - | |
| CE2 | - | |
| HE2 | - | |
| CZ | - | |
| HZ | - | |
| HB# | 3.07 | |
| CD# | 132 | |
| HD# | 7.22 | |
| CE# | 131.6 | |
| HE# | 7.33 | |
| 96 | Val | 1096 |
| N | 122.7 | - |
| HN | 7.9 | - |
| CA | 62.2 | 62.2 |
| HA | 4.01 | 4.03 |
| CO | 175.2 | 175.3 |
| CB | 33 | 33 |
| HB | 1.95 | 1.96 |
| CG1 | - | - |
| HG1# | 0.87 | - |
| CG2 | - | - |
| HG2# | 0.84 | - |
| CG# | 21 | 21 |
| HG# | 0.86 | 0.87 |
| 97 | Arg | 1097 |
| N | 126.2 | 126.2 |
| HN | 8.18 | 8.21 |
| CA | 55.9 | 55.8 |
| HA | 4.34 | 4.34 |
| CO | 174.5 | 174.5 |
| CB | 30.9 | 30.7 |
| HB1 | 1.75 | 1.74 |
| HB2 | 1.64 | 1.64 |
| CG | 27.4 | - |
| HG1 | 1.63 | - |
| HG2 | 1.58 | 1.59 |
| CD | 43.7 | - |
| HD1 | - | - |
| HD2 | - | - |
| NE | - | - |
| HE | - | - |
| NH1 | - | - |
| HH11 | - | - |
| HH12 | - | - |
| NH2 | - | - |
| HH21 | - | - |
| HH22 | - | - |
| HB# | - | - |
| HG# | 1.6 | - |
| HD# | 3.16 | 3.15 |
| HH1# | - | - |
| HH2# | - | - |
| NH# | - | - |
| HH# | - | - |
| 98 | Ile | 1098 |
| N | 121.6 | 121.7 |
| HN | 7.92 | 7.98 |
| CA | 59.8 | 59.6 |
| HA | 4.37 | 4.36 |
| CO | 176.2 | 176.1 |
| CB | 41.2 | 41.2 |
| HB | 1.67 | 1.71 |
| CG1 | 27.1 | 27.1 |
| HG11 | 1.38 | 1.38 |
| HG12 | 1.03 | 1.03 |
| CG2 | 17.2 | 17.5 |
| HG2# | 0.85 | 0.87 |
| CD1 | 13.7 | 13.3 |
| HD1# | 0.84 | 0.82 |
| HG1# | - | - |
| 99 | Cys | 1099 |
| N | 124.4 | 125.1 |
| HN | 7.65 | 8.26 |
| CA | 56.7 | 56.3 |
| HA | 3.5 | 3.52 |
| CO | 174.1 | 174.1 |
| CB | 43.8 | 43.5 |
| HB1 | 2.44 | 2.46 |
| HB2 | 1.8 | 2 |
| HB# | - | - |
| 100 | Ser | 1100 |
| N | 114.4 | 115 |
| HN | 8.34 | 8.35 |
| CA | 59.2 | 59.1 |
| HA | 4.36 | 4.37 |
| CO | 178.2 | 178 |
| CB | 64.5 | 64.4 |
| HB1 | - | - |
| HB2 | - | - |

| | | |
|---|---|---|
| HG | - | - |
| HB# | 4.02 | 4.02 |
| | | |
| 101 | Lys | 1101 |
| | | |
| N | 127.5 | 127.2 |
| HN | 9.08 | 9.04 |
| CA | 58.3 | 58.2 |
| HA | 4.28 | 4.27 |
| CO | 178.3 | 178.2 |
| CB | 31.6 | 31.7 |
| HB1 | 2.12 | 2.1 |
| HB2 | 1.94 | 1.97 |
| CG | 26.1 | 25.8 |
| HG1 | 1.73 | 1.66 |
| HG2 | 1.5 | 1.57 |
| CD | 29.3 | 29.4 |
| HD1 | 1.76 | 1.77 |
| HD2 | 1.68 | - |
| CE | 42.3 | 42.3 |
| HE1 | - | - |
| HE2 | - | - |
| NZ | - | - |
| HZ# | - | - |
| HB# | - | - |
| HG# | - | - |
| HD# | - | - |
| HE# | 3.01 | 3.04 |
| | | |
| 102 | Ser | Ser |
| | | |
| N | 119 | 119 |
| HN | 8.75 | 8.75 |
| CA | 60.9 | - |
| HA | 4.29 | - |
| CO | 175.4 | - |
| CB | 63.1 | - |
| HB1 | - | - |
| HB2 | - | - |
| HG | - | - |
| HB# | 4.05 | - |
| | | |
| 103 | Tyr | |
| | | |
| N | 123.1 | |
| HN | 7.14 | |
| CA | 60.4 | |
| HA | 3.97 | |
| CO | 173.1 | |
| CB | 38.5 | |
| HB1 | 2.44 | |
| HB2 | 2.22 | |
| CD1 | - | |
| HD1 | - | |
| CD2 | - | |
| HD2 | - | |
| CE1 | - | |
| HE1 | - | |
| CE2 | - | |
| HE2 | - | |
| HH | - | |
| HB# | - | |
| CD# | 131.2 | |
| HD# | 6.45 | |
| CE# | 117.7 | |
| HE# | 6.73 | |
| | | |
| 104 | Leu | |
| | | |
| N | 112.1 | |
| HN | 7.34 | |
| CA | 55.4 | |
| HA | 4.18 | |
| CO | 176.2 | |
| CB | 40.7 | |
| HB1 | 1.88 | |
| HB2 | 1.83 | |
| CG | 28.6 | |
| HG | 1.72 | |
| CD1 | 25.3 | |
| HD1# | 1.04 | |
| CD2 | 23.4 | |
| HD2# | 0.85 | |
| HB# | 1.85 | |
| CD# | - | |
| HD# | - | |
| | | |
| 105 | Thr | 1105 |
| | | |
| N | 111.6 | - |
| HN | 7.29 | - |
| CA | 60.7 | 60.6 |
| HA | 4.24 | - |
| CO | 171.3 | - |
| CB | 71.5 | 71.5 |
| HB | 3.93 | - |
| CG2 | 21.3 | - |
| HG1 | - | - |
| HG2# | 1.08 | - |
| | | |
| 106 | Leu | 1106 |
| | | |
| N | 125.1 | 125.3 |
| HN | 8.46 | 8.47 |
| CA | 53.3 | 53.2 |
| HA | 4.39 | - |
| CO | 174.7 | - |
| CB | 45.7 | 45.6 |
| HB1 | 1.34 | - |
| HB2 | 0.95 | - |
| CG | 26.5 | - |
| HG | 1.09 | - |
| CD1 | 27.4 | - |
| HD1# | 0.64 | - |
| CD2 | 24 | - |
| HD2# | 0.48 | - |
| HB# | - | - |
| CD# | - | - |
| HD# | - | - |
| | | |
| 107 | Glu | 1107 |
| | | |
| N | 130.2 | - |
| HN | 8.7 | - |
| CA | 58 | - |
| HA | 3.89 | - |
| CO | 177.3 | - |
| CB | 28.9 | - |
| HB1 | 1.84 | - |
| HB2 | 1.78 | - |
| CG | 35.9 | - |
| HG1 | - | - |
| HG2 | - | - |
| CD | - | - |
| HB# | 1.81 | - |
| HG# | 2.21 | - |
| | | |
| 108 | Asn | 1108 |
| | | |
| N | 118.2 | - |
| HN | 8.95 | - |
| CA | 55.6 | - |
| HA | 3.76 | - |
| CO | 172.5 | - |
| CB | 36.3 | - |
| HB1 | 1.98 | - |
| HB2 | 1.08 | - |
| CG | - | - |
| ND2 | 116.5 | - |
| HD21 | 7.44 | - |
| HD22 | 6.81 | - |
| HB# | - | - |
| HD2# | - | - |
| | | |
| 109 | Gly | 1109 |
| | | |
| N | 103.3 | - |
| HN | 7.56 | - |
| CA | 46.4 | - |
| HA1 | 4.34 | - |
| HA2 | 3.88 | - |
| CO | 171.9 | - |
| HA# | - | - |
| | | |
| 110 | Lys | 1110 |

| | | |
|---|---|---|
| N | 121 | - |
| HN | 9.01 | - |
| CA | 54.8 | 54.8 |
| HA | 4.6 | 4.6 |
| CO | 173.5 | 173.5 |
| CB | 37.7 | 37.6 |
| HB1 | - | - |
| HB2 | - | - |
| CG | 24.9 | 24.9 |
| HG1 | 1.32 | - |
| HG2 | 1.24 | - |
| CD | 29.2 | 29.1 |
| HD1 | 1.69 | - |
| HD2 | 1.61 | - |
| CE | 42.4 | 42.4 |
| HE1 | - | - |
| HE2 | - | - |
| NZ | - | - |
| HZ# | - | - |
| HB# | 1.51 | 1.51 |
| HG# | - | - |
| HD# | - | - |
| HE# | 2.95 | 2.95 |
| | | |
| 111 | Val | 1111 |
| | | |
| N | 119.5 | 119.5 |
| HN | 7.83 | 7.8 |
| CA | 61 | 60.9 |
| HA | 4.54 | 4.57 |
| CO | 175.5 | 175.4 |
| CB | 34.4 | 34.4 |
| HB | 1.62 | 1.61 |
| CG1 | 21.7 | - |
| HG1# | 0.74 | 0.74 |
| CG2 | 21.4 | - |
| HG2# | 0.61 | 0.61 |
| CG# | 21.6 | 21.6 |
| HG# | - | - |
| | | |
| 112 | Phe | 1112 |
| | | |
| N | 127.8 | 127.5 |
| HN | 8.97 | 8.88 |
| CA | 56.5 | 56.5 |
| HA | 4.75 | 4.76 |
| CO | 174.7 | 174.4 |
| CB | 39.9 | 40.3 |
| HB1 | 3.05 | 3.04 |
| HB2 | 2.9 | 2.88 |
| CD1 | - | - |
| HD1 | - | - |
| CD2 | - | - |
| HD2 | - | - |
| CE1 | - | - |
| HE1 | - | - |
| CE2 | - | - |
| HE2 | - | - |
| CZ | 129.4 | - |
| HZ | 7.2 | - |
| HB# | - | - |
| CD# | 131.9 | - |
| HD# | 7.15 | - |
| CE# | 131 | - |
| HE# | 7.17 | - |
| | | |
| 113 | Leu | 1113 |
| | | |
| N | 130.6 | 130.1 |
| HN | 8.76 | 8.69 |
| CA | 54.5 | 54.6 |
| HA | 5.05 | 5.13 |
| CO | 176 | 176.1 |
| CB | 45 | 44.9 |
| HB1 | 1.88 | 1.84 |
| HB2 | 1.29 | 1.31 |
| CG | 27.6 | 27.7 |
| HG | 1.35 | 1.48 |
| CD1 | 26.7 | 26.2 |
| HD1# | 0.94 | 0.89 |
| CD2 | 24.1 | 24.9 |
| HD2# | 0.88 | 0.89 |
| HB# | - | - |
| CD# | - | - |
| HD# | 0.91 | 0.89 |
| | | |
| 114 | Thr | 1114 |
| | | |
| N | 124.4 | 121.2 |
| HN | 9.37 | 9.19 |
| CA | 62.1 | 61.3 |
| HA | 4.56 | 4.64 |
| CO | 174 | 173.7 |
| CB | 70.6 | 71.5 |
| HB | 4.07 | 4.09 |
| CG2 | 21.4 | 21.6 |
| HG1 | - | - |
| HG2# | 1.26 | 1.19 |
| | | |
| 115 | Gly | 1115 |
| | | |
| N | 113.5 | 111.3 |
| HN | 8.9 | 8.42 |
| CA | 45.9 | 45.6 |
| HA1 | 4.15 | 4.17 |
| HA2 | 3.89 | 3.89 |
| CO | 175.4 | 173.3 |
| HA# | - | - |
| | | |
| 116 | Gly | 1116 |
| | | |
| N | 106.6 | 106.1 |
| HN | 8.45 | 7.92 |
| CA | 46.7 | 45 |
| HA1 | - | 3.85 |
| HA2 | - | 3.79 |
| CO | 173.6 | 172.7 |
| HA# | 3.77 | 3.82 |
| | | |
| 117 | Asp | 1117 |
| | | |
| N | 116.9 | 120.7 |
| HN | 8.5 | 8.19 |
| CA | 53.3 | 52.7 |
| HA | 4.61 | 4.92 |
| CO | 176.3 | 175.4 |
| CB | 40.2 | 41.3 |
| HB1 | - | 2.76 |
| HB2 | - | 2.54 |
| CG | - | - |
| HB# | 2.69 | - |
| | | |
| 118 | Leu | 1118 |
| | | |
| N | 124.1 | 122.1 |
| HN | 7.73 | 8.14 |
| CA | 53.7 | 54.3 |
| HA | 4.45 | 4.19 |
| CO | - | 176.3 |
| CB | 40.7 | 40.9 |
| HB1 | 1.6 | 1.66 |
| HB2 | 1.47 | 1.4 |
| CG | 24.8 | 24.8 |
| HG | 1.42 | 1.39 |
| CD1 | - | 25.5 |
| HD1# | - | 0.93 |
| CD2 | - | 23.1 |
| HD2# | - | 0.85 |
| HB# | 1.49 | - |
| CD# | 22.5 | - |
| HD# | 0.9 | - |
| | | |
| 119 | Pro | 1119 |
| | | |
| N | - | - |
| CA | 64.3 | 63.7 |
| HA | 4.28 | 4.35 |
| CO | 178 | 176.4 |
| CB | 31.4 | 34.9 |
| HB1 | 2.26 | 2.36 |
| HB2 | 1.93 | 2.16 |
| CG | 27.8 | 24.6 |
| HG1 | 2.18 | 1.97 |
| HG2 | 2.05 | 1.79 |

| | | |
|---|---|---|
| CD | 50.7 | 50.5 |
| HD1 | 3.9 | 3.65 |
| HD2 | 3.7 | 3.48 |
| HB# | - | - |
| HG# | - | - |
| HD# | - | - |
| | | |
| 120 | Ala | 1120 |
| | | |
| N | 125.6 | 127 |
| HN | 8.89 | 8.31 |
| CA | 53.6 | 54 |
| HA | 3.93 | 4.3 |
| CO | 176 | 177.3 |
| CB | 18 | 19.3 |
| HB# | 1.39 | 1.47 |
| | | |
| 121 | Leu | 1121 |
| | | |
| N | 112.1 | 115.8 |
| HN | 8.18 | 8.2 |
| CA | 56.5 | 55.2 |
| HA | 3.57 | 3.96 |
| CO | 175.2 | 175.9 |
| CB | 40 | 41.9 |
| HB1 | 2.18 | 1.83 |
| HB2 | 1.55 | 1.48 |
| CG | 27.2 | 27.1 |
| HG | 1.56 | 1.55 |
| CD1 | 26.2 | 26.2 |
| HD1# | 0.92 | 0.91 |
| CD2 | 23.9 | 23.7 |
| HD2# | 0.9 | 0.76 |
| HB# | - | - |
| CD# | - | - |
| HD# | 0.91 | - |
| | | |
| 122 | Asp | 1122 |
| | | |
| N | 120.4 | 119.1 |
| HN | 7.59 | 7.52 |
| CA | 57.3 | 57.3 |
| HA | 4.05 | 4.07 |
| CO | 176.3 | 177.5 |
| CB | 40.9 | 41.2 |
| HB1 | 2.78 | - |
| HB2 | 2.5 | - |
| CG | - | - |
| HB# | - | 2.55 |
| | | |
| 123 | Gly | 1123 |
| | | |
| N | 116.3 | 115.6 |
| HN | 8.57 | 8.41 |
| CA | 45.1 | 45.5 |
| HA1 | 4.33 | 4.21 |
| HA2 | 3.63 | 3.71 |
| CO | 174.9 | 174.1 |
| HA# | - | - |
| | | |
| 124 | Ala | 1124 |
| | | |
| N | 125.6 | 123.6 |
| HN | 8.44 | 8.26 |
| CA | 52.9 | 52.6 |
| HA | 4.43 | 4.37 |
| CO | 175.5 | 175.4 |
| CB | 20.4 | 20.8 |
| HB# | 1.4 | 1.34 |
| | | |
| 125 | Arg | 1125 |
| | | |
| N | 117.4 | 118.8 |
| HN | 8.49 | 8.71 |
| CA | 54.7 | 54.7 |
| HA | 5.51 | 5.42 |
| CO | 175.1 | 175.5 |
| CB | 34.2 | 33.4 |
| HB1 | 1.94 | 1.9 |
| HB2 | 1.74 | 1.77 |
| CG | 27.8 | 27.9 |
| HG1 | 1.73 | 1.73 |
| HG2 | 1.65 | 1.63 |
| CD | 44 | 44 |
| HD1 | - | - |
| HD2 | - | - |
| NE | 105.6 | - |
| HE | 7.19 | - |
| NH1 | - | - |
| HH11 | - | - |
| HH12 | - | - |
| NH2 | - | - |
| HH21 | - | - |
| HH22 | - | - |
| HB# | - | - |
| HG# | - | - |
| HD# | 3.18 | 3.17 |
| HH1# | - | - |
| HH2# | - | - |
| NH# | - | - |
| HH# | - | - |
| | | |
| 126 | Val | 1126 |
| | | |
| N | 116.3 | 116.7 |
| HN | 8.9 | 8.9 |
| CA | 58.5 | 58.3 |
| HA | 5.24 | 5.3 |
| CO | 174.6 | 174.6 |
| CB | 35.2 | 35 |
| HB | 1.23 | 1.27 |
| CG1 | 23.6 | 23.4 |
| HG1# | 0.5 | 0.51 |
| CG2 | 21.4 | 21.4 |
| HG2# | 0.33 | 0.35 |
| CG# | - | - |
| HG# | - | - |
| | | |
| 127 | Glu | 1127 |
| | | |
| N | 122.9 | 122.9 |
| HN | 8.2 | 8.27 |
| CA | 54 | 53.9 |
| HA | 5.13 | 5.12 |
| CO | 174.6 | 174.6 |
| CB | 33 | 33 |
| HB1 | 2.07 | 2.06 |
| HB2 | 1.9 | 1.89 |
| CG | 36.3 | 36.3 |
| HG1 | 2.36 | 2.35 |
| HG2 | 2.13 | 2.13 |
| CD | - | - |
| HB# | - | - |
| HG# | - | - |
| | | |
| 128 | Phe | 1128 |
| | | |
| N | 124.1 | 124.1 |
| HN | 8.83 | 8.78 |
| CA | 57.1 | 57.1 |
| HA | 5.03 | 5.01 |
| CO | 175.2 | 175.2 |
| CB | 42.7 | 42.7 |
| HB1 | 2.74 | 2.74 |
| HB2 | 2.59 | 2.6 |
| CD1 | - | - |
| HD1 | - | - |
| CD2 | - | - |
| HD2 | - | - |
| CE1 | - | - |
| HE1 | - | - |
| CE2 | - | - |
| HE2 | - | - |
| CZ | 130.1 | - |
| HZ | 6.84 | - |
| HB# | - | - |
| CD# | 132.6 | - |
| HD# | 6.97 | - |
| CE# | 131.2 | - |
| HE# | 7.09 | - |
| | | |
| 129 | Arg | 1129 |
| | | |
| N | 120.4 | 120.4 |
| HN | 8.6 | 8.62 |

| | | |
|---|---|---|
| CA | 54.8 | 54.8 |
| HA | 4.67 | 4.67 |
| CO | 175.2 | - |
| CB | 34.4 | 34.4 |
| HB1 | 1.89 | 1.88 |
| HB2 | 1.75 | 1.74 |
| CG | 27.2 | - |
| HG1 | 1.72 | - |
| HG2 | 1.62 | - |
| CD | 43.7 | - |
| HD1 | 3.28 | - |
| HD2 | 3.12 | - |
| NE | 105.4 | - |
| HE | 7.1 | - |
| NH1 | - | - |
| HH11 | - | - |
| HH12 | - | - |
| NH2 | - | - |
| HH21 | - | - |
| HH22 | - | - |
| HB# | - | - |
| HG# | - | - |
| HD# | - | - |
| HH1# | - | - |
| HH2# | - | - |
| NH# | - | - |
| HH# | - | - |

| 130 | Cys |
|---|---|
| N | 121.4 |
| HN | 8.99 |
| CA | 55.2 |
| HA | 5.04 |
| CO | 174.3 |
| CB | 41.5 |
| HB1 | 3.06 |
| HB2 | 2.56 |
| HB# | - |

| 131 | Asp |
|---|---|
| N | 125.5 |
| HN | 8.98 |
| CA | 54 |
| HA | 4.7 |
| CO | 173.3 |
| CB | 38 |
| HB1 | 2.9 |
| HB2 | 2.4 |
| CG | - |
| HB# | - |

| 132 | Pro |
|---|---|
| N | - |
| CA | 65.3 |
| HA | 4.44 |
| CO | 176.9 |
| CB | 32 |
| HB1 | 2.4 |
| HB2 | 1.89 |
| CG | 28 |
| HG1 | 2.16 |
| HG2 | 2.07 |
| CD | 50.5 |
| HD1 | 3.84 |
| HD2 | 3.65 |
| HB# | - |
| HG# | - |
| HD# | - |

| 133 | Asp |
|---|---|
| N | 114.9 |
| HN | 8.57 |
| CA | 56 |
| HA | 4.42 |
| CO | 173.5 |
| CB | 38.7 |
| HB1 | 2.89 |
| HB2 | 2.81 |
| CG | - |
| HB# | 2.83 |

| 134 | Phe |
|---|---|
| N | 117.8 |
| HN | 8.32 |
| CA | 56.8 |
| HA | 4.93 |
| CO | 175.1 |
| CB | 43.4 |
| HB1 | 3.23 |
| HB2 | 2.52 |
| CD1 | - |
| HD1 | - |
| CD2 | - |
| HD2 | - |
| CE1 | - |
| HE1 | - |
| CE2 | - |
| HE2 | - |
| CZ | 129.9 |
| HZ | 7.29 |
| HB# | - |
| CD# | 131.5 |
| HD# | 6.8 |
| CE# | 131.6 |
| HE# | 7.27 |

| 135 | His |
|---|---|
| N | 114 |
| HN | 9.18 |
| CA | 53.9 |
| HA | 5.12 |
| CO | 172.7 |
| CB | 30.9 |
| HB1 | 3.3 |
| HB2 | 3.27 |
| ND1 | - |
| HD1 | - |
| CD2 | 120.2 |
| HD2 | 7 |
| CE1 | - |
| HE1 | - |
| NE2 | - |
| HE2 | - |
| HB# | 3.28 |

| 136 | Leu |
|---|---|
| N | 126.2 |
| HN | 8.72 |
| CA | 56.5 |
| HA | 4.26 |
| CO | 176.6 |
| CB | 43.2 |
| HB1 | 1.74 |
| HB2 | 1.46 |
| CG | 27.4 |
| HG | 1.15 |
| CD1 | 26.4 |
| HD1# | 0.63 |
| CD2 | 25.5 |
| HD2# | 0.54 |
| HB# | - |
| CD# | - |
| HD# | 0.58 |

| 137 | Val |
|---|---|
| N | 130.1 |
| HN | 8.92 |
| CA | 62.6 |
| HA | 4.13 |
| CO | 174.7 |
| CB | 32.3 |
| HB | 1.17 |
| CG1 | - |
| HG1# | 0.86 |
| CG2 | - |
| HG2# | 0.8 |
| CG# | 20.9 |

| | | |
|---|---|---|
| HG# | - | |
| | | |
| 138 | Gly | |
| | | |
| N | 115.7 | |
| HN | 8.39 | |
| CA | 42.9 | |
| HA1 | 4.55 | |
| HA2 | 3.6 | |
| CO | 172.7 | |
| HA# | - | |
| | | |
| 139 | Ser | |
| | | |
| N | 115.8 | |
| HN | 8.63 | |
| CA | 58.7 | |
| HA | 4.48 | |
| CO | 174.5 | |
| CB | 64.2 | |
| HB1 | 4.08 | |
| HB2 | 3.91 | |
| HG | - | |
| HB# | - | |
| | | |
| 140 | Ser | |
| | | |
| N | 117.4 | |
| HN | 8.5 | |
| CA | 60.1 | |
| HA | 4.45 | |
| CO | 173.4 | |
| CB | 64.5 | |
| HB1 | 4.38 | |
| HB2 | 3.96 | |
| HG | - | |
| HB# | - | |
| | | |
| 141 | Arg | 1141 |
| | | |
| N | 119.3 | - |
| HN | 7.67 | 7.67 |
| CA | 54.8 | 54.8 |
| HA | 5.57 | 5.6 |
| CO | 175.4 | - |
| CB | 34.6 | - |
| HB1 | 1.9 | - |
| HB2 | 1.77 | - |
| CG | 27.3 | - |
| HG1 | 1.69 | - |
| HG2 | 1.61 | - |
| CD | 43.8 | - |
| HD1 | - | - |
| HD2 | - | - |
| NE | 105.9 | - |
| HE | 7.34 | - |
| NH1 | - | - |
| HH11 | - | - |
| HH12 | - | - |
| NH2 | - | - |
| HH21 | - | - |
| HH22 | - | - |
| HB# | 1.85 | - |
| HG# | - | - |
| HD# | 3.17 | - |
| HH1# | - | - |
| HH2# | - | - |
| NH# | - | - |
| HH# | - | - |
| | | |
| 142 | Ser | 1142 |
| | | |
| N | 117.3 | - |
| HN | 8.59 | - |
| CA | 59.1 | 59 |
| HA | 4.74 | 4.72 |
| CO | 173 | 173 |
| CB | 67.2 | 67 |
| HB1 | 3.85 | - |
| HB2 | 3.8 | - |
| HG | - | - |
| HB# | 3.82 | 3.82 |
| | | |
| 143 | Val | 1143 |
| | | |
| N | 124.1 | 123.8 |
| HN | 8.94 | 8.91 |
| CA | 60.8 | 60.7 |
| HA | 5.3 | 5.37 |
| CO | 174.9 | 175 |
| CB | 35.8 | 35.8 |
| HB | 2.04 | 2.04 |
| CG1 | 21.4 | 21.4 |
| HG1# | 0.97 | 0.97 |
| CG2 | 20.6 | 20.6 |
| HG2# | 0.92 | 0.91 |
| CG# | - | - |
| HG# | 0.94 | - |
| | | |
| 144 | Cys | 1144 |
| | | |
| N | 130 | 129.9 |
| HN | 8.79 | 8.8 |
| CA | 54.8 | 54.8 |
| HA | 4.26 | 4.27 |
| CO | 174.2 | 174.1 |
| CB | 39.3 | 39.4 |
| HB1 | 2.72 | 2.54 |
| HB2 | 0.65 | 0.63 |
| HB# | - | - |
| | | |
| 145 | Ser | 1145 |
| | | |
| N | 124.6 | 124.6 |
| HN | 8.4 | 8.42 |
| CA | 56.9 | 56.9 |
| HA | 4.77 | 4.72 |
| CO | 174.5 | 174.4 |
| CB | 64.5 | 64.4 |
| HB1 | 3.73 | 3.72 |
| HB2 | 3.63 | 3.63 |
| HG | - | - |
| HB# | - | - |
| | | |
| 146 | Gln | 1146 |
| | | |
| N | 126.4 | 126 |
| HN | 9.4 | 9.41 |
| CA | 56.7 | 56.7 |
| HA | 3.75 | 3.77 |
| CO | 176.3 | 176.3 |
| CB | 26.9 | 26.8 |
| HB1 | 2.17 | 2.17 |
| HB2 | 2.03 | 2.04 |
| CG | 34.3 | 34.2 |
| HG1 | 2.36 | 2.36 |
| HG2 | 2.29 | 2.3 |
| CD | - | - |
| NE2 | 112.5 | 112.5 |
| HE21 | 7.46 | 7.42 |
| HE22 | 6.7 | 6.73 |
| HB# | - | - |
| HG# | 2.32 | 2.32 |
| HE2# | - | - |
| | | |
| 147 | Gly | 1147 |
| | | |
| N | 104.5 | 104.5 |
| HN | 8.18 | 8.32 |
| CA | 46.2 | 46.2 |
| HA1 | 3.77 | 3.77 |
| HA2 | 3.22 | 3.21 |
| CO | 172.7 | - |
| HA# | - | - |
| | | |
| 148 | Gln | 1148 |
| | | |
| N | 118.7 | 118.2 |
| HN | 7.53 | 7.5 |
| CA | 53.7 | 53.7 |
| HA | 4.5 | 4.49 |
| CO | 175.5 | 175.3 |
| CB | 32.6 | 32.6 |
| HB1 | 2.02 | 2.01 |
| HB2 | 1.85 | 1.85 |

| | | |
|---|---|---|
| CG | 33.6 | 33.6 |
| HG1 | - | - |
| HG2 | - | - |
| CD | - | - |
| NE2 | 112.5 | - |
| HE21 | 7.53 | - |
| HE22 | 6.79 | - |
| HB# | - | - |
| HG# | 2.27 | 2.26 |
| HE2# | - | - |
| | | |
| 149 | Trp | 1149 |
| | | |
| N | 127.1 | 126.7 |
| HN | 8.37 | 8.37 |
| CA | 57 | 56.9 |
| HA | 5.13 | 5.13 |
| CO | 179.7 | 179.7 |
| CB | 30.9 | 30.9 |
| HB1 | 3.41 | 3.4 |
| HB2 | 2.62 | 2.62 |
| CD1 | 127.1 | 127.2 |
| HD1 | 6.4 | 6.43 |
| NE1 | 130.5 | 129.9 |
| HE1 | 10.22 | 10.09 |
| CE3 | 120.8 | - |
| HE3 | 7.26 | - |
| CZ2 | 113.2 | 113.3 |
| HZ2 | 6.92 | 6.97 |
| CZ3 | 122.8 | - |
| HZ3 | 6.82 | - |
| CH2 | 122.9 | - |
| HH2 | 6.63 | 6.63 |
| HB# | - | - |
| | | |
| 150 | Ser | 1150 |
| | | |
| N | 117.8 | - |
| HN | 9.24 | 9.23 |
| CA | 60.8 | - |
| HA | 4.29 | - |
| CO | 174.7 | - |
| CB | 63.6 | - |
| HB1 | - | - |
| HB2 | - | - |
| HG | - | - |
| HB# | 4.3 | - |
| | | |
| 151 | Thr | |
| | | |
| N | 111.9 | |
| HN | 7.62 | |
| CA | 58.9 | |
| HA | 5.19 | |
| CO | 172.7 | |
| CB | 70.8 | |
| HB | 4.53 | |
| CG2 | 22.3 | |
| HG1 | - | |
| HG2# | 1.49 | |
| | | |
| 152 | Pro | 1152 |
| | | |
| N | - | - |
| CA | 62.3 | 62.3 |
| HA | 4.68 | 4.68 |
| CO | 176.4 | 176.4 |
| CB | 32.4 | 32.3 |
| HB1 | 2.45 | 2.45 |
| HB2 | 1.78 | 1.77 |
| CG | 27.6 | 27.6 |
| HG1 | 2.1 | - |
| HG2 | 2.04 | - |
| CD | 51.5 | 51.5 |
| HD1 | 3.99 | 3.98 |
| HD2 | 3.73 | 3.72 |
| HB# | - | - |
| HG# | 2.08 | 2.08 |
| HD# | - | - |
| | | |
| 153 | Lys | 1153 |
| | | |
| N | 129.6 | 129.5 |
| HN | 8.37 | 8.35 |
| CA | 54.7 | 54.6 |
| HA | 3.44 | 3.42 |
| CO | 174.7 | 174.7 |
| CB | 33.1 | 33.1 |
| HB1 | 1.21 | 1.2 |
| HB2 | 0.53 | 0.53 |
| CG | 24.3 | - |
| HG1 | 1.11 | - |
| HG2 | 0.36 | - |
| CD | 30.1 | - |
| HD1 | 1.2 | - |
| HD2 | 0.99 | - |
| CE | 42.8 | - |
| HE1 | 2.86 | - |
| HE2 | 2.75 | - |
| NZ | - | - |
| HZ# | - | - |
| HB# | - | - |
| HG# | - | - |
| HD# | - | - |
| HE# | - | - |
| | | |
| 154 | Pro | |
| | | |
| N | - | - |
| CA | 63.1 | - |
| HA | 4.72 | - |
| CO | 173.6 | - |
| CB | 33.5 | - |
| HB1 | 1.88 | - |
| HB2 | 1.42 | - |
| CG | 26.6 | - |
| HG1 | 1.49 | - |
| HG2 | 0.38 | - |
| CD | 49.7 | - |
| HD1 | 2.9 | - |
| HD2 | 1.81 | - |
| HB# | - | - |
| HG# | - | - |
| HD# | - | - |
| | | |
| 155 | His | |
| | | |
| N | 109.2 | - |
| HN | 7.9 | - |
| CA | 54.4 | - |
| HA | 4.71 | - |
| CO | 172.4 | - |
| CB | 31.8 | - |
| HB1 | 3.07 | - |
| HB2 | 3.03 | - |
| ND1 | - | - |
| HD1 | - | - |
| CD2 | 120.1 | - |
| HD2 | 7.05 | - |
| CE1 | - | - |
| HE1 | - | - |
| NE2 | - | - |
| HE2 | - | - |
| HB# | 3.05 | - |
| | | |
| 156 | Cys | 1156 |
| | | |
| N | 121.4 | - |
| HN | 9.31 | - |
| CA | 54.6 | - |
| HA | 5.07 | 5.06 |
| CO | 173.6 | - |
| CB | 40.9 | - |
| HB1 | 2.67 | 2.67 |
| HB2 | 2.55 | 2.54 |
| HB# | - | - |
| | | |
| 157 | Gln | 1157 |
| | | |
| N | 126.8 | 127.2 |
| HN | 9.06 | 9.06 |
| CA | 53.7 | - |
| HA | 4.84 | 4.84 |
| CO | 174.6 | - |
| CB | 32.3 | - |

| | | |
|---|---|---|
| HB1 | 1.68 | 1.68 |
| HB2 | 1.32 | 1.32 |
| CG | 33.3 | - |
| HG1 | - | - |
| HG2 | - | - |
| CD | - | - |
| NE2 | 113.3 | - |
| HE21 | 7.72 | - |
| HE22 | 6.92 | - |
| HB# | - | - |
| HG# | 2.45 | - |
| HE2# | - | - |

| | |
|---|---|
| 158 | Val |
| N | 127.8 |
| HN | 8.57 |
| CA | 63.2 |
| HA | 3.46 |
| CO | 174.7 |
| CB | 32.3 |
| HB | 1.86 |
| CG1 | 20.4 |
| HG1# | 0.85 |
| CG2 | 21.3 |
| HG2# | 0.77 |
| CG# | - |
| HG# | 0.8 |
| 159 | Asn |
| N | 129.2 |
| HN | 7.93 |
| CA | 55.2 |
| HA | 4.33 |
| CO | 180.5 |
| CB | 38 |
| HB1 | 2.77 |
| HB2 | 2.6 |
| CG | - |
| ND2 | 110.4 |
| HD21 | 7.51 |
| HD22 | 6.6 |
| HB# | - |
| HD2# | - |

# APPENDIX D: *CNS* scripts for structure calculation

## D-1 *wrapper* script

```
!@NODE_NAME:

{=====>} evaluate ($fileroot = "")

{=====>} evaluate ($start_count=1)
{=====>} evaluate ($end_count=10)

{=====>} evaluate ($seed = 1879)

parameter
  reset
  evaluate ($par_nonbonded = PROLSQ) {* set forecefield to use *}
  @@PROTOCOLS:../PARALLHDG_5.1/parallhdg5.1.pro
end

struct reset @@../mtf/M2.mtf end

!------------------------------------------------------------------------------
! Read experimental restraints

noe
  reset
  nrestraints = 100000              ! allocate space for NOEs
  ceiling 100
end

noe
  set echo off message off end

   class c  @../restraints/restraint_files_061201/or_c-noesy.tbl
   class n  @../restraints/restraint_files_061201/or_NOESY_HSQC.tbl

  set echo on message on end


  monomers  * 2
  averaging * sum
  potential * soft
  scale     * 1.0
  sqconstant * 1.0
  sqexponent * 2
  soexponent * 1
  rswitch   * 1.0
  sqoffset  * 0.0
  asymptote * 2.0

  ?
end



! phi angle restraints

coupling
  reset
  class = scthmsqcha
  coefficients 6.98 -1.38 1.72 -60
  potential=square
  force = 1.0        .
  set echo off message off end

  @@../restraints/avujhnha.tbl

  set echo on message on end
end


!------------------------------------------------------------------------------

print thresh=10000.0 noe
print thresh=10000.0 cdih


!------------------------------------------------------------------------------
! NCS restraints

ncs restraints
 initialize
```

```
 group
  equi ((resid 92:111) or (resid 128:159))
  equi ((resid 1092:1111) or (resid 1128:1159))
  weight = 0.05
 end
 ?
end

!-------------------------------------------------------------------------------
! Append node name (if any set) to $fileroot

if ($exist_node_name = TRUE) then
   evaluate ($fileroot = $fileroot + "_" + $node_name)
end if

!-------------------------------------------------------------------------------
! Multiply $seed by $node_id (if set)

if ($exist_node_id = TRUE) then
   evaluate ($seed = $seed * $node_id)
end if

inline @PROTOCOLS:rand.cns

inline @PROTOCOLS:rrsa_swap.cns

coupling
   force = 1.0
end

inline @PROTOCOLS:refine_swap.cns

stop
```

## D-2 *rand.cns* script

```
remarks rand_partial.inp
remarks Reference:
remarks   Nilges, M. (1994)
remarks Modified for 4-D ambiguity by Andrew Raine 14/2/95


if ($exist_start_count = FALSE) then
   {=====>} evaluate ($start_count=1)       {* number of structures *}
end if

if ($exist_end_count = FALSE) then
   {=====>} evaluate ($end_count=10)        {* number of structures *}
end if

if ($exist_seed= FALSE) then
   evaluate ($seed = 1872)
end if

evaluate ($seed  = $seed * ($start_count + 1 ))

set seed $seed end

!-----------------------------------------------------

evaluate ($init_t = 2000 )
evaluate ($bath = $init_t)


  do (refx = -15.0) (resid 92:159)
  do (refx = 15.0) (resid 1092:1159)
  do (refy = 0) (all)
  do (refz = 0) (all)

evaluate ($kharm = 0.0001)
  do (harm = $kharm) (all)

 do (mass=100) (all)

evaluate ($count = $start_count)

while ($count <= $end_count ) loop main

   {* 0 ========================================== initial values *}
    do (x = (random()-0.5)*20) (all)
    do (y = (random()-0.5)*20) (all)
    do (z = (random()-0.5)*20) (all)

!#######################################
! this bit spreads the monomers out
!#######################################
coor
  orient selection=(resid 92:159)
end

! translate A along x axis so A & B end up a bit spread after rotation below
coor
  translate vector=(-1,0,0) distance=15.0
end

! copy coordinates from A to B
set message off echo off end
 identify (store9) (resid 92:159)
 show min (store9) (resid 92:159)
evaluate ($acounter = $result)
 show max (store9) (resid 92:159)
evaluate ($aend = $result)
evaluate ($natoms = $aend - $acounter + 1)
 do (store9=0) (all) !reset store9

while ($acounter le $aend) loop counter
  evaluate ($bcounter = $acounter + $natoms) !assume segid B follows A
    show element (x) (id $acounter)
    do (x = $result) (id $bcounter)
    show element (y) (id $acounter)
    do (y = $result) (id $bcounter)
    show element (z) (id $acounter)
    do (z = $result) (id $bcounter)
  evaluate ($acounter = $acounter +1)
end loop counter

! translate B along x axis so A & B end up a bit spread out
coor
  translate selection=(resid 1092:1159) vector=(1,0,0) distance=30.0
end
```

```
set message on echo on end
!#############################################

   parameter nbonds
       nbxmod 4
       wmin 0.01
       cutnb 100
       tolerance 45
       repel=1.54
       rexp=4
       irexp=1
       rcon=4.0
   end end

   evaluate ($knoe = 0.5)
   noe scale * $knoe end
   evaluate ($kcoup = $knoe/25)
   coupling force=$kcoup end

   evaluate ($kbon = 0.0005 )  ! bonds   increased ARCR 21/12/95
   evaluate ($kang = 0.0005 )  ! angles
   evaluate ($kvdw = 0.1)  ! vdw

   evaluate ($kimp = 0.0)  ! impropers (added by ARCR, 'cos MN missed it out)
   evaluate ($kdihe = 0.0)  ! diherals (added by BOS, for PARALLHDG5.1)

   {* 1 ================================================ minimization *}

!---------------------------------
! Set up selection (in store1) for heavy atoms in "reduced representation"
@PROTOCOLS:setup_reduced.cns

   igroup
     interaction (not store1) (all)
     weights bond $kbon angl $kang impr $kimp dihe $kdihe vdw 0 elec 0 end
     interaction (store1) (store1)
     weights bond $kbon angl $kang impr $kimp dihe $kdihe vdw $kvdw end
     interaction (resid 92:159) (resid 1092:1159)
     weights vdw 0 end
   end

   flags exclude * include vdw {harm} end

   minimize powell nstep 50 end

   flags exclude * include bonds angle {impr} {dihe} vdw noe coup ncs {harm} end

   {* 2 ================================================ high temperature dynamics *}

   do (vx = maxwell($init_t)) (all)
   do (vy = maxwell($init_t)) (all)
   do (vz = maxwell($init_t)) (all)

   do (fbeta=10) (all)
  evaluate ($timestep = 0.03)
  evaluate ($nstep = 300)

   while ($kbon < 0.01) loop stage1

     evaluate ($kbon = min(0.25, $kbon * 1.25))
     evaluate ($kang = $kbon)

     igroup
       interaction (not store1) (all)
       weights bond $kbon angl $kang impr $kimp dihe $kdihe vdw 0 elec 0 end
       interaction (store1) (store1)
       weights bond $kbon angl $kang impr $kimp dihe $kdihe vdw $kvdw end
       interaction (resid 92:159) (resid 1092:1159)
       weights vdw 0 end
     end

     dynamics  cartesian
       nstep=$nstep   timestep=$timestep    tcoupling=true
       temperature=$bath  nprint=50
     end

   end loop stage1

   {* 4 ================================================ now with all atoms *}

   flags exclude * include bonds angle {impr} {dihe} vdw noe coup ncs {harm} end

   parameter nbonds
     nbxmod -3
     cutnb=4.5
```

```
      tolerance=0.5
      repel=1.28
    end end

    evaluate ($knoe = 5.0) ! ARCR + PJD
    noe scale * $knoe end
    evaluate ($kcoup = $knoe/25)
    coupling force=$kcoup end

    igroup
      interaction (all) (all)
      weights bond 0.02 angl 0.02 impr 0 dihe 0 vdw 0.002 elec 0 end
        interaction (resid 92:159) (resid 1092:1159)
        weights vdw 0 end
    end

    do (fbeta=10.0) (all)

    evaluate ($bath = 1500)

    dynamics cartesian
      nstep=500 timestep=0.003
      tcoupling=true
      temperature=$bath nprint=50
    end

    igroup
      interaction (all) (all)
      weights bond 0.05 angl 0.05 impr 0 dihe 0 vdw 0.005 elec 0 end
        interaction (resid 92:159) (resid 1092:1159)
        weights vdw 0 end
    end
    dynamics cartesian
      nstep=500 timestep=0.003   tcoupling=true
      temperature=$bath nprint=50
    end

    igroup
      interaction (all) (all)
      weights bond 0.1 angl 0.1 impr 0 dihe 0 vdw 0.01 elec 0 end
        interaction (resid 92:159) (resid 1092:1159)
        weights vdw 0 end
    end

    do (fbeta=10.0) (all)

    dynamics cartesian
      nstep=500 timestep=0.002   tcoupling=true
      temperature=$bath nprint=50
    end

    remarks seed = $seed

    evaluate ($filename = "rand_" + $fileroot + "_" + encode($count) + ".pdb")
    write coordinates output =$filename end

    evaluate ($count = $count + 1)
  end loop main

  ! Belt and braces effort to make sure that all the nbonds stuff is right for
  ! the next stage

  parameter
    nbonds
      nbxmod=-2          {* initial value for nbxmod - modified later *}
      repel=0.64         {* initial value for repel - modified later *}
    end
  end
```

## D-3 *rrsa_swap.cns* script

```
REMARKS rrsa_partial.inp
REMARKS DGSA.inp from the x-plor manual.

if ($exist_start_count = FALSE) then
   {=====>} evaluate ($start_count=0)       {* number of structures *}
end if

if ($exist_end_count = FALSE) then
   {=====>} evaluate ($end_count=5)         {* number of structures *}
end if

if ($exist_seed= FALSE) then
   evaluate ($seed = 8764)
end if

evaluate ($seed  = $seed * ($start_count + 1))

{* Doubled high_steps and cool_steps 21/8/95 *}

{=====>} evaluate ($init_t = 2000 )    {* initial simulated annealing temp *}
{=====>} evaluate ($high_steps = 4000 ) {* total number of steps at high temp *}
                                {* includes "loop anne" with swapping *}
{=====>} evaluate ($cool_steps = 2000 ) {* total number of steps during cooling *}

parameter
   nbonds
     nbxmod=-2
     wmin=0.01
     cutnb=4.5
     tolerance=0.5
     repel=0.64              {* initial value for repel - modified later *}
     rexp=4
     irexp=1
     rcon=4.
   end
end

noe
      averaging  * sum
      potential  * soft !BOS
      scale      * 25   !PJD
      sqconstant * 1.0
      asymptote  * 1.0  !BOS
            rswitch    * 1.0  !BOS

            ?
end

! coupling constant restraints

coupling
   force = 1.0
end


evaluate ($count = $start_count)
while ($count <= $end_count ) loop main

{=====>} evaluate ($nreassign = 0)     {* number of times velocity reassigned *}
{=====>} evaluate ($bath = $init_t)    {* reset temperature *}

  evaluate ($filename="rand_" + $fileroot + "_" + encode($count)+".pdb")

{* test for the correct enantiomer                    *}
{=====>}       {* if you want to by-pass this test because the substructures *}
          {* were tested previously, simply remove the -1 from the next *}
          {* statement                                *}
  for $image in ( 1 -1 ) loop imag
    coor initialize end
    coor @@$filename
    do (x=x * $image) ( known )
    identify (store1) (not known)        {* set store1 to unknowns *}

    {* 1 ================== create local ideal geometry by template fitting *}
    {*              this takes also care of unknown atoms         *}

    set message=off echo=off end
    coor copy end          {* store current coordinates in comparison set *}

{=====>}                {* the user has to supply a template coordinate set *}
    coor @@../template/M2_template.pdb
    for $id in id ( tag ) loop fit             {* loop over residue tags *}
```

```
    coordinates                {* lsq fitting using known coordinates *}
      fit select = ( byresidue (id $id) and not store1 )
    end
                               {* store fitted template coordinates for this residue *}
   coor copy selection=( byresidue (id $id) ) end

  end loop fit

  coor swap end

  set message=on echo=on end



  {* 2 ================================== minimization of bonds, vdw and NOEs *}
  coupling force=0.2 end
  parameter  nbonds  nbxmod=-2 repel=0.64   end  end
  igroup interaction (all) (all) weights * 1.  vdw 20. end end

  flags exclude * include {bond} vdw noe coup ncs end
  energy end
  while ($ncs > 50.0) loop symm
   flags exclude * include {bond} vdw {noe} {coup} ncs end
   minimize powell nstep=400 nprint=20 end
   flags exclude * include {bond} vdw noe coup ncs end
   minimize powell nstep=400 nprint=20 end
  end loop symm

  flags exclude * include bond vdw noe coup ncs end
  minimize powell nstep=400 nprint=20 end

  {* 3 ================================================================ include angles *}

  flags include angl end

  minimize powell nstep=400 nprint=20 end

  {* 4 ============== dynamics, slowly introducing chirality and planarity *}

  evaluate ($nstep1 = int($high_steps/16))
  evaluate ($nstep2 = int($high_steps/8))

  flags include impr dihe end

  do (mass=100) (all)
  do (fbeta=10.0) (all)

  igroup inter (all) (all) weights * 0.1 impr 0.005 dihe 0.005 vdw 20. end end
  dynamics  cartesian
    nstep=$nstep1 time=0.003
    tcoup=true temperature=$bath nprint=100
  end

  igroup inter (all) (all) weights * 0.1 impr 0.05 dihe 0.05 vdw 20. end end
  dynamics  cartesian
    nstep=$nstep1 time=0.003
    tcoup=true temperature=$bath nprint=100
  end

  igroup inter (all) (all) weights * 0.2 impr 0.1 dihe 0.1 vdw 20. end end
  dynamics  cartesian
    nstep=$nstep1 time=0.003
    tcoup=true temperature=$bath nprint=100
  end

  parameter  nbonds repel=1.28   end  end
  igroup inter (all) (all) weights * 0.2 impr 0.2 dihe 0.2 vdw 0.01 end end
  dynamics  cartesian
    nstep=$nstep1 time=0.003
    tcoup=true temperature=$bath nprint=100
  end

  parameter nbonds nbxmod=-3  end  end
  igroup inter (all) (all) weights * 0.4 impr 0.4 dihe 0.4 vdw 0.003 end end
  dynamics  cartesian
    nstep=$nstep2 time=0.003
    tcoup=true temperature=$bath nprint=100
  end

  igroup inter (all) (all) weights * 1.0 impr 1.0 dihe 1.0 vdw 0.003 end end
  dynamics  cartesian
    nstep=$nstep2 time=0.003
    tcoup=true temperature=$bath nprint=100
  end

  if ($image = 1) then
```

```
      do (store7=x) ( all ){ *store first image in stores *}
      do (store8=y) ( all )
      do (store9=z) ( all )
      do (store4=vx) ( all )
      do (store5=vy) ( all )
      do (store6=vz) ( all )
    end if

  end loop imag

  {* 5 ========== establish the correct handedness of the structure *}

  energy end
  evaluate ($e_minus=$ener)
  coor copy end
  do (x=store7) ( all )
  do (y=store8) ( all )
  do (z=store9) ( all )
  energy end
  evaluate ($e_plus=$ener)
  if ( $e_plus > $e_minus ) then
    evaluate ($hand=-1 )
    coor swap end
  else
    evaluate ($hand= 1 )
    do (vx=store4) ( all )
    do (vy=store5) ( all )
    do (vz=store6) ( all )
  end if


 display $cpu

 { Here is where we do some dynamics with the prochiral centers swapping }
   { define prochiral centres to be swapped & floated }
   { store2 is set AND ids are stored in bcomp & qcomp}
   @PROTOCOLS:setup_swap.cns
    (
     swap_sel=&swap_sel;
    )

   @PROTOCOLS:float_prochirals.cns

 display $cpu


   evaluate ($nstep1=int($high_steps / 20.0 ) )

   do (vx = maxwell($bath)) (all)
   do (vy = maxwell($bath)) (all)
   do (vz = maxwell($bath)) (all)

   evaluate ($ini_noe = 5.0)        evaluate ($fin_noe = 50.0)
   evaluate ($k_noe = $ini_noe)
   evaluate ($noe_fac = ($fin_noe/$ini_noe)^(1/10.0))

   evaluate ($hightemp_count = 0)

   while ($hightemp_count < 10) loop anne
    evaluate ($hightemp_count = $hightemp_count + 1)
    evaluate ($k_noe = $k_noe * $noe_fac)

    noe scale * $k_noe end

 display $cpu

    { Do swaps }
    @PROTOCOLS:swap_metropolis.cns
     (
       toswap = methyl amido {methylene};
       swap_sel=&swap_sel;
       bath=$bath;        {needs $bath}
       swap_stats=$swap_stats;
     )
 display $cpu

    flags
     exclude *
     include angl bond impr dihe vdw noe coup ncs
    end

    evaluate ($k_vdw = $k_noe * 0.002)
    igroup interaction (all) (all) weights * 1. vdw $k_vdw end end


    dynamics  cartesian
```

```
      nstep=$nstep1   timestep=0.003
      tcoupling=true  temperature=$bath  nprint=50
      ntrfrq=9999
    end

    evaluate ($critical=$temp/$bath)
    if ($critical > 2.0) then
      evaluate ($nreassign = $nreassign + 1)
      mini powell nstep 100 nprint 10 end
      do (vx=maxwell($bath)) ( all )
      do (vy=maxwell($bath)) ( all )
      do (vz=maxwell($bath)) ( all )
    end if

  end loop anne

{stop}

    {* 6 ======= ========================== increase vdw interaction and cool *}

  coupling force=1.0 end

  evaluate ($final_t = 100)    { K }
  evaluate ($tempstep = 50)    { K }

  evaluate ($ncycle = ($init_t-$final_t)/$tempstep)
  evaluate ($nstep = int($cool_steps/$ncycle))

  evaluate ($ini_rad  = 1.28)      evaluate ($fin_rad  = 1.0)
  evaluate ($ini_con = 0.012)      evaluate ($fin_con= 16.0)
  evaluate ($ini_flt = 25.0)       evaluate ($fin_flt = 500.0)
  evaluate ($k_flt = $ini_flt)
  evaluate ($flt_fac = ($fin_flt/$ini_flt)^(1/$ncycle))
  evaluate ($bath  = $init_t)
  evaluate ($k_vdw = $ini_con)
  evaluate ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))
  evaluate ($radius=   $ini_rad)
  evaluate ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))

  evaluate ($i_cool = 1)
  while ($i_cool < $ncycle) loop cool
    evaluate ($i_cool=$i_cool+1)

    evaluate ($bath  = $bath  - $tempstep)
    evaluate ($k_vdw=min($fin_con,$k_vdw*$k_vdwfact))
    evaluate ($radius=max($fin_rad,$radius*$radfact))
    evaluate ($k_flt=min($fin_flt,$k_flt*$flt_fac))
    parameter  nbonds repel=$radius rcon=$k_vdw end end

    parameter
      angle (store3) (all) (store3) $k_flt TOKEN
      angle (all)    (all) (store3) $k_flt TOKEN
    end

    igroup interaction (all) (all) weights * 1. vdw $k_vdw end end

    do (fbeta=10.0) (all)

    { Do swaps }
    @PROTOCOLS:swap_metropolis.cns
     (
       toswap = methyl amido {methylene};
       swap_sel=&swap_sel;
       bath=$bath;        {needs $bath}
       swap_stats=$swap_stats;
     )
    flags
      exclude *
      include angl bond impr dihe vdw noe coup ncs
    end

    dynamics  cartesian
      nstep=$nstep time=0.003
      tcoup=true temperature=$bath nprint=$nstep
    end


{====>}                              {* abort condition *}

  evaluate ($critical=$temp/$bath)
  if ($critical > 2.0) then
    evaluate ($nreassign = $nreassign + 1)
    mini powell nstep 100 nprint 10 end
    do (vx=maxwell($bath)) ( all )
    do (vy=maxwell($bath)) ( all )
    do (vz=maxwell($bath)) ( all )
```

```
    end if

  end loop cool

  {* 7 ==========---========== final minimization *}

{ Do swaps }
display $cpu

@PROTOCOLS:swap_metropolis.cns
 (
      toswap = methyl amido methylene;
      swap_sel=&swap_sel;
      bath=$bath;       {needs $bath}
      swap_stats=$swap_stats;
 )
display $cpu

flags
 exclude *
 include angl bond impr dihe vdw noe coup ncs
end

  minimize powell nstep= 400 nprint=50 end

{ lash up to avoid ids overflowing b & q fields of pdb }
  do (b = 0.0) (all)
  do (q = 1.0) (all)

  print thresh=0.5 noe
  evaluate ($rms_noe=$result)
  evaluate ($violations_noe=$violations)

  coupling
   print threshold=1.5 all
   evaluate ($rms_coup=$result)
   evaluate ($violations_coup=$violations)
  end

  evaluate ($filename="rrsa_" + $fileroot + "_" + encode($count)+".pdb")

  remarks final noe energy $noe
  remarks final coup energy $coup
  remarks violations noe: $violations_noe coup: $violations_coup
  remarks seed = $seed
  remarks velocity reassigned $nreassign times
  remarks swaps tried immediately before minimization: $swap_stats.tries
  remarks swaps rejected: $swap_stats.rejectpc percent

  write coordinates output =$filename end

  evaluate ($count=$count+1)
end loop main

! What status are we in?

@PROTOCOLS:status.cns
```

## D-4 *refine.cns* script

```
REMARK refine_partial.inp
REMARK refinement protocol with ambiguous restraints
REMARK Nilges, M. (1993).Proteins 17, 297-309.


if ($exist_start_count = FALSE) then
   {====>} evaluate ($start_count=0)       {* number of structures *}
end if

if ($exist_end_count = FALSE) then
   {====>} evaluate ($end_count=5)         {* number of structures *}
end if

if ($exist_seed= FALSE) then
   evaluate ($seed = 8764)
end if

evaluate ($seed  = $seed * ($start_count + 1))

{* Doubled high_steps and cool_steps 21/8/95 *}

evaluate ($high_steps  = 10000)
evaluate ($cool_steps_1 = 10000)
evaluate ($cool_steps_2 = 4000)
evaluate ($init_t  = 2000.01)

!------------------------------------------------------------------------
! set the weights for the experimental energy terms

evaluate ($knoe  = 25.0)  ! noes
evaluate ($asym  = 0.1)  ! slope of NOE potential
evaluate ($kcdi  = 0.0)  ! torsion angles


noe
  averaging  * sum
  potential  * soft
  scale      * 25
  sqconstant * 1.0
  sqexponent * 2
  soexponent * 1
  rswitch    * 0.5
  sqoffset   * 0.0
  asymptote  * 0.1

  ?
end

! phi angle restraints

coupling
   !force = 1.0
end


evaluate ($rcon  = 0.003)

flags
  exclude *
  include angl bond impr vdw noe coup dihe ncs
end

evaluate ($count = $start_count)

while ($count <= $end_count) loop stru
  evaluate ($filename="rrsa_" + $fileroot + "_" + encode($count)+".pdb")

  coor init end
  coor @@$filename

evaluate ($bath = $init_t)
do (mass = 100.0) (all)
do (fbeta = 10.0) (all)

{ define prochiral centres to be swapped & floated }
{ store2 is set AND ids are stored in bcomp & qcomp}
@PROTOCOLS:setup_swap.cns
 (
   swap_sel=&swap_sel;
 )

@PROTOCOLS:float_prochirals.cns

  {* ================================ high temperature stage (reduced non-bonded) *}
```

```
  evaluate ($knoe  = 25.0)  ! noes
  evaluate ($asym  = 0.1)  ! slope of NOE potential
  noe
    averaging  * sum
    potential  * soft
    scale      * $knoe
    sqconstant * 1.0
    sqexponent * 2
    soexponent * 1
    rswitch    * 0.5
    sqoffset   * 0.0
    asymptote  * $asym
  end

  parameters
    nbonds
      nbxmod 3
      wmin  = 0.01  ! warning off
      cutnb  = 100  ! nonbonded cutoff
      tolerance 45
      repel  = 1.54  ! scale factor for vdW radii = 1 ( L-J radii)
      rexp  = 4    ! exponents in (r^irex - R0^irex)^rexp
      irex  = 1
      rcon  = 4.0  ! actually set the vdW weight
    end
  end

inline @PROTOCOLS:setup_reduced.cns

  igroup
    interaction (not store1) (all)
    weights * 1 angl 0.4 impr 0.1 dihe 0.1 vdw 0 elec 0 end
    interaction (store1) (store1)
    weights * 1 angl 0.4 impr 0.1 dihe 0.1 vdw 0.1 end
  end

  dynamics  cartesian
    nstep=$high_steps
    timestep=0.005


    tcoupling = true
    temperature = $bath
    nprint=50

    ntrfr = 99999999
  end

  parameters { starting conditions for cooling }
    nbonds
      atom
      nbxmod 5
      wmin  = 0.01  ! warning off
      cutnb  = 7.0  ! nonbonded cutoff
      tolerance 0.5
      repel  = 1.28  ! scale factor for vdW radii = 1 ( L-J radii)
      rexp  = 4    ! exponents in (r^irex - R0^irex)^rexp
      irex  = 1
      rcon  = 4.0  ! actually set the vdW weight
    end
  end

  {* ================================================================ first cooling stage *}

  evaluate ($final_t = 1000)    { K }
  evaluate ($tempstep =  50)    { K }

  evaluate ($ncycle = ($init_t-$final_t)/$tempstep)
  evaluate ($nstep = int($cool_steps_1/$ncycle))

  evaluate ($ini_rad = 1.28)       evaluate ($fin_rad = 1.0)
  evaluate ($ini_con = 0.012)      evaluate ($fin_con = 16.0)
  evaluate ($ini_ang = 0.25)       evaluate ($fin_ang = 1.0)
  evaluate ($ini_imp = 0.1)        evaluate ($fin_imp = 1.0)
  evaluate ($ini_dihe = 0.1)       evaluate ($fin_dihe = 1.0)
  evaluate ($ini_asy = 0.1)        evaluate ($fin_asy = 1.0)
  evaluate ($ini_flt = 25.0)       evaluate ($fin_flt = 500.0)

  evaluate ($k_vdw = $ini_con)
  evaluate ($k_vdwfact = ($fin_con/$ini_con)^(1/$ncycle))
  evaluate ($radius=   $ini_rad)
  evaluate ($radfact = ($fin_rad/$ini_rad)^(1/$ncycle))
  evaluate ($k_ang = $ini_ang)
  evaluate ($ang_fac = ($fin_ang/$ini_ang)^(1/$ncycle))
  evaluate ($k_imp = $ini_imp)
```

```
  evaluate ($imp_fac = ($fin_imp/$ini_imp)^(1/$ncycle))
  evaluate ($k_dihe = $ini_dihe)
  evaluate ($dihe_fac = ($fin_dihe/$ini_dihe)^(1/$ncycle))
  evaluate ($asym = $ini_asy)
  evaluate ($asy_fac = ($fin_asy/$ini_asy)^(1/$ncycle))
  evaluate ($k_flt = $ini_flt)
  evaluate ($flt_fac = ($fin_flt/$ini_flt)^(1/$ncycle))

  do (vx = maxwell($bath)) (all)
  do (vy = maxwell($bath)) (all)
  do (vz = maxwell($bath)) (all)

  evaluate ($i_cool = 1)
  while ($i_cool < $ncycle) loop cool1
    evaluate ($i_cool=$i_cool+1)

    evaluate ($bath  = $bath  - $tempstep)
    evaluate ($k_vdw=min($fin_con,$k_vdw*$k_vdwfact))
    evaluate ($radius=max($fin_rad,$radius*$radfact))
    evaluate ($k_ang = $k_ang*$ang_fac)
    evaluate ($k_imp = $k_imp*$imp_fac)
    evaluate ($k_dihe = $k_dihe*$dihe_fac)
    evaluate ($asym  = $asym*$asy_fac)
    evaluate ($k_flt = $k_flt*$flt_fac)

    igroup interaction (all) (all) weights
      * 1 angles $k_ang improper $k_imp dihedral $k_dihe
    end end
    parameter nbonds
      cutnb=7.0 rcon=$k_vdw nbxmod=5 repel=$radius
    end      end
    parameter
     angle (store3) (all) (store3) $k_flt TOKEN
     angle (all)    (all) (store3) $k_flt TOKEN
    end
    noe
       asymptote * $asym
    end

    { Do swaps }
    @PROTOCOLS:swap_metropolis.cns
     (
      toswap = methyl amido methylene;
      swap_sel=&swap_sel;
      bath=$bath;        {needs $bath}
      swap_stats=$swap_stats;
     )
    flags
     exclude *
     include angl bond impr vdw noe coup dihe ncs
    end

    dynamics  cartesian
      nstep=$nstep  timestep=0.003
      tcoupling = true  temperature = $bath   nprint=$nstep
      ntrfr = 99999999
    end
  end loop cool1

{* ========================================================== second cooling stage *}

evaluate ($final_t = 100)     { K }
evaluate ($tempstep = 50)     { K }

evaluate ($ncycle = ($bath-$final_t)/$tempstep)
evaluate ($nstep = int($cool_steps_2/$ncycle))

evaluate ($i_cool = 1)
while ($i_cool < $ncycle) loop cool2
  evaluate ($i_cool=$i_cool+1)

  evaluate ($bath  = $bath  - $tempstep)

  { Do swaps }
  @PROTOCOLS:swap_metropolis.cns
   (
    toswap = methyl amido methylene;
    swap_sel=&swap_sel;
    bath=$bath;        {needs $bath}
    swap_stats=$swap_stats;
   )
  flags
    exclude *
    include angl bond impr vdw noe coup dihe ncs
  end
  dynamics  cartesian
```

```
      nstep=$nstep  timestep=0.003
      tcoupling = true  temperature = $bath   nprint=$nstep
      ntrfr = 99999999
    end
  end loop cool2

  evaluate ($stats = 2 * $nstep)

  dynamics  cartesian
      nstep=$stats  timestep=0.003
      tcoupling = true  temperature = $bath   nprint=$nstep
      ntrfr = 99999999
  end

{* ========== Powell energy minimization ========== *}
{ Do swaps }
@PROTOCOLS:swap_downhill.cns
 (
  toswap = methyl amido methylene;
  swap_sel=&swap_sel;
  bath=$bath;              {needs $bath}
  swap_stats=$swap_stats;
 )
flags
 exclude *
 include angl bond impr vdw noe coup dihe ncs
end

  mini powell nstep= 500 nprint= 50 end

{ Do swaps }
@PROTOCOLS:swap_downhill.cns
 (
  toswap = methyl amido methylene;
  swap_sel=&swap_sel;
  bath=$bath;              {needs $bath}
  swap_stats=$swap_stats;
 )
flags
 exclude *
 include angl bond impr vdw noe coup dihe ncs
end

  mini powell nstep= 500 nprint= 50 end

{ Do swaps }
@PROTOCOLS:swap_downhill.cns
 (
  toswap = methyl amido methylene;
  swap_sel=&swap_sel;
  bath=$bath;              {needs $bath}
  swap_stats=$swap_stats;
 )
flags
 exclude *
 include angl bond impr vdw noe coup dihe ncs
end

  mini powell nstep= 500 nprint= 50 end

{ lash up to avoid ids overflowing b & q fields of pdb and producing havoc in PROCHECK with q=0.0}
  do (b = 0.0) (all)
  do (q = 1.0) (all)

 evaluate ($violfile = $fileroot + encode($count) + ".viol")
 set print $violfile end

 print threshold=0.5 noe
 evaluate ($rms_noe=$result)
 evaluate ($violations_noe=$violations)
 coupling
  print threshold=1.2 all
  evaluate ($rms_coup=$result)
  evaluate ($violations_coup=$violations)
 end
 print thres=0.018 bonds      {6*sigma Engh & Huber}
 evaluate ($rms_bond=$result)
 print thres=3.04 angles      {6*sigma Engh & Huber}
 evaluate ($rms_angl=$result)
 print thres=5. impropers
 evaluate ($rms_impr=$result)
 print thres=5. dihedrals
 evaluate ($rms_dihe=$result)

 close $violfile end
 set print_file=OUTPUT end
```

```
  remarks ============================================================
  remarks          overall,bonds,angles,improper,vdw,noe,coup,dihe,ncs
  remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $coup, $dihe, $ncs
  remarks ============================================================
  remarks          bonds,angles,impropers,noe,coup,dihe,ncs
  remarks rms-d: $rms_bond,$rms_angl,$rms_impr,$rms_noe,$rms_coup,$rms_dihe,$rms_ncs
  remarks ============================================================
  remarks            noe,  coup
  remarks violations.: $violations_noe, $violations_coup
  remarks ============================================================
  remarks seed = $seed

  evaluate ($file = "refine_" + $fileroot + "_" + encode($count) + ".pdb")
  write coor output= $file end

  evaluate ($count = $count + 1)
end loop stru
```

## APPENDIX E: *Aria*-based scripts for structure analysis

Note: the *anal_NOESY_HSQC.inp* script, the *filt_NOESY_HSQC.inp* script are equivalent to the anal_c-noesy.inp, and the filt_c-noesy.inp script respectively; but applies to the $^{15}$N-edited noesy experiment.

### E-1 *analyse* script

```
setenv PROTOCOLS ../../../protocols/swapping+J

cns_solve < anal_c-noesy.inp > anal_c-noesy.log
cns_solve < anal_NOESY_HSQC.inp > anal_NOESY_HSQC.log
cns_solve < anal_J.inp > anal_J.log
```

### E-1.1 *anal_c-noesy.inp* script

```
! Script to analyse ensemble of structures and the ambiguous NOE used
! to generate them in order to produce new restraints for the next
! iteration of Michael Nilges' ARIA approach
! received from Andrew Raine March, 1997

!read parameter file
parameter
  reset
  @CNS_TOPPAR:protein-allhdg.param
end

!read psf
struct
  reset
  @@../../../mtf/M2.mtf
end

!------------------------------------------------------------------------------
! Set up equivalences for methyls and aromatics etc.

  set echo off message off end
    inline @define_methyls.cns
  set echo on message on end

! To get around a bug in the noe analyse statements, we need to
! look at all the restraint files individually (actually, we could do them
! all at the same time if we were prepared to put them all into one class)


!------------------------------------------------------------------------------
! Read experimental restraints

noe
  reset
  nrestraints = 50000              ! allocate space for NOEs
  ceiling 100
end

noe
!   reset
!   nrestraints = 50000 { allocate space for NOEs }
!   ceiling    = 1000 { maximum force constant that is allowed }

  set message off echo off end
  class      c @@../../../restraints/restraint_files_041201/or_c-noesy.tbl

  set message on echo on end
  averaging   c sum
  monomers    c 2
  potential   c soft
  scale       c 1 { initial scale }
  sqoffset    c 0.0
  sqconstant  c 1.0
  sqexponent  c 2
  soexponent  c 1
  asymptote   c 2.0
  rswitch     c 1.0

 set echo on message on end

 averaging  * sum
 potential  * soft
 scale      * 1.0
```

```
    sqconstant * 1.0
    sqexponent * 2
    soexponent * 1
    rswitch    * 1.0
    sqoffset   * 0.0
    asymptote  * 2.0
    ?
  end

  aria                                  { Initialize the new analysis features of Michael's}
    analyze
      reset
      mode intra
    end

    countviol reset end
  end

  for $file in (
                @files.cns
        ) loop main


  !evaluate ($count = 0)
  !evaluate ($end_count=30)
  !
  !while ($count < $end_count)  loop main
  !     evaluate ($count=$count+1)
  !evaluate ($file="refine__"+encode($count)+".pdb")

    coord init end
    coord @@$file

    aria
      analyse               { Accumulate info about all the restraints over }
              !mini         { all the structures.  Each restraint will be associated }
              aver          { with either the minimum or mean value over the ensemble }
            end
    end

    aria
      countviol
            threshold 0.3
          end
    end

    !print threshold 0.3 noe

  end loop main

  {*************************************************************************}
  {This section for an analysis of each violated restraint (in the log file)}
  aria countviol
    list * 0.0 {fraction of ensemble in which restraints are violated}
                        { 0.0 means list all restraints violated at all, 1.0 means}
                        { all restraints violated in ALL structures }

  end end

  {*************************************************************************}
  {This section for table of nuclei contributing to each restraint}
  set print ../contributions/contrib_c.txt end

  aria analyse          { report on the stats gathered earlier }
     minnumber 1
     maxnumber 1000
     cutoff 10.0      { distance - nothing further than this}
     level  1.0       { fraction contribution - list all possibilities making}
               {                 up to this fraction of the intensity}
     sort *          { Lists all possibilities for each restraint which     }
               { satisfy the criteria above, in order of contribution }
                                   { to total intensity (ISPA!) }
  end end

  {
  {*************************************************************************}
  {This section for new restraint table excluding violated restraints and nuclei}
  {contributing less than a set fraction of the intensity of the crosspeak}
  set print ../restraints/c.tbl end

  aria
    countviol
      exclude * 1.0 { fraction of ensemble in which restraints are violated}
              { 0.0 means exclude ANY violations                }
              { 0.01 < x < 0.99 means exclude restraints violated in }
              {               more than this fraction of ensemble }
```

```
  end

  analyse          { report on the stats gathered earlier }
   minnumber 1
   maxnumber 1000
   cutoff 10.0      { distance }
   level  0.99      { fraction contribution}
   or-restraint *     { Creates new .tbl from the restraints, taking into }
                                    { account the exclusions due to violations and the }
                                    { limited assignments due to the other criteria }
  end
end
}

stop
_
```

## E-1.1 *anal_J.inp* script

```
! Script to analyse ensemble of structures and the coupling constants used
! iteration of Michael Nilges' ARIA approach
! received from Andrew Raine March, 1997

!read parameter file
parameter
  reset
  @CNS_TOPPAR:protein-allhdg.param
end

!read psf
struct
  reset
  @@../../../mtf/M2.mtf
end

!------------------------------------------------------------------------------
! Set up equivalences for methyls and aromatics etc.

  set echo off message off end
    inline @define_methyls.cns
  set echo on message on end

! To get around a bug in the noe analyse statements, we need to
! look at all the restraint files individually (actually, we could do them
! all at the same time if we were prepared to put them all into one class)


!------------------------------------------------------------------------------
! Read experimental restraints

noe
  reset
  nrestraints = 50000              ! allocate space for NOEs
  ceiling 100
end

coupling
  reset
  class = SCT
  coefficients 6.98 -1.38 1.72 -60
!  coefficients 6.51 -1.76 1.60 -60
  potential square
  force 1.0

!  set echo off message off end
  @@../../../restraints/avujhnha.tbl
!  set echo on message on end

end

for $file in (
                @files.cns
        ) loop main
  evaluate ($count = $file - "../pdbs/refine__" - ".pdb")
  evaluate ($count = decode($count))


!evaluate ($count = 0)
!evaluate ($end_count=30)
!
!while ($count < $end_count)  loop main
!      evaluate ($count=$count+1)
!evaluate ($file="../pdbs/refine__"+encode($count)+".pdb")

  evaluate ($Jfile = "../pdbs/" + encode($count) + ".J")
  evaluate ($phifile = "../pdbs/" + encode($count) + ".phi")

  coord init end
  coord @@$file

  set print_file = $Jfile end
  coupling
            print threshold 1.0 all
            !print threshold 0.0 all
  end
  close $Jfile end
  set print_file = OUTPUT end

  set display = $phifile end
  display resid              phi
  display N  N
  evaluate ($resi = 92)
  while ($resi < 1159) loop phi
```

```
    evaluate ($resim1 = $resi - 1)
    pick dihedral (resid $resim1 and name C) (resid $resi and name N)
                (resid $resi and name CA) (resid $resi and name C) geometry
    evaluate ($phi.$count.$resi = $result)
    display $resi               $phi.$count.$resi
    evaluate ($resi = $resi + 1)
  end loop phi

end loop main


stop
-
```

## E-2 *filter* script & *filt_c-noesy.inp* script

### E-2.1 *filter* script

```
setenv PROTOCOLS ../../../protocols/swapping+J

cns_solve < filt_c-noesy.inp > filt_c-noesy.log
cns_solve < filt_NOESY_HSQC.inp > filt_NOESY_HSQC.log
```

### E-2.2 *filt_c-noesy.inp* script

```
! Script to analyse ensemble of structures and the ambiguous NOE used
! to generate them in order to produce new restraints for the next
! iteration of Michael Nilges' ARIA approach
! received from Andrew Raine March, 1997

!read parameter file
parameter
  reset
  @CNS_TOPPAR:protein-allhdg.param
end

!read psf
struct
  reset
  @@../../../mtf/M2.mtf
end

!------------------------------------------------------------------------
! Set up equivalences for methyls and aromatics etc.

  set echo off message off end
    inline @define_methyls.cns
  set echo on message on end

! To get around a bug in the noe analyse statements, we need to
! look at all the restraint files individually (actually, we could do them
! all at the same time if we were prepared to put them all into one class)


!------------------------------------------------------------------------
! Read experimental restraints

noe
  reset
  nrestraints = 50000              ! allocate space for NOEs
  ceiling 100
end

noe
!   reset
!   nrestraints = 50000 { allocate space for NOEs }
!   ceiling    = 1000 { maximum force constant that is allowed }

  set message off echo off end
  class      c @@../../../restraints/restraint_files_250801/or_c-noesy.tbl


  set message on echo on end
  averaging   c sum
  monomers    c 2
  potential   c soft
  scale       c 1 { initial scale }
  sqoffset    c 0.0
  sqconstant  c 1.0
  sqexponent  c 2
  soexponent  c 1
  asymptote   c 2.0
  rswitch     c 1.0


  set echo on message on end

  averaging  * sum
  potential  * soft
  scale      * 1.0
  sqconstant * 1.0
  sqexponent * 2
  soexponent * 1
  rswitch    * 1.0
  sqoffset   * 0.0
  asymptote  * 2.0
```

```
    ?
  end

  aria
    analyze
      reset
      mode intra
    end

    countviol reset end
  end

  for $file in (
                  @files.cns
          ) loop main


  !evaluate ($count = 0)
  !evaluate ($end_count=30)
  !
  !while ($count < $end_count)  loop main
  !     evaluate ($count=$count+1)
  !evaluate ($file="refine__"+encode($count)+".pdb")

    coord init end
    coord @@$file

    aria
      analyse   { Accumulate info about all the restraints over }
                    { all the structures.  Each restraint will be associated }
                !minimum      { with either the minimum }
                average      { or mean }
                !accumulate   { or r6sum value over the ensemble }
              end
    end

    aria
      countviol
                threshold 0.5
              end
    end

    !print threshold 0.5 noe

  end loop main

  {
  {*************************************************************************}
  {This section for an analysis of each violated restraint (in the log file)}
  aria countviol
    list * 0.0 {fraction of ensemble in which restraints are violated}
                          { 0.0 means list all restraints violated at all, 1.0 means}
                          { all restraints violated in ALL structures }

  end end
  }

  {
  {*************************************************************************}
  {This section for table of nuclei contributing to each restraint}
  set print ../contributions/contrib_c-noesy.txt end

  aria analyse          { report on the stats gathered earlier }
      minnumber 1
      maxnumber 1000
      cutoff 10.0       { distance - nothing further than this}
      level  1.0        { fraction contribution - list all possibilities making}
                {               up to this fraction of the intensity}
      sort *            { Lists all possibilities for each restraint which    }
                { satisfy the criteria above, in order of contribution }
                                        { to total intensity (ISPA!) }
  end end
  }

  {*************************************************************************}
  {This section for new restraint table excluding violated restraints and nuclei}
  {contributing less than a set fraction of the intensity of the crosspeak}
  set print ../restraints/filt_c-noesy.tbl end

  aria
    countviol
      exclude * 1.0 { fraction of ensemble in which restraints are violated}
              { 0.0 means exclude ANY violations                }
              { 0.01 < x < 0.99 means exclude restraints violated in }
              {               more than this fraction of ensemble }
    end
```

```
  analyse          { report on the stats gathered earlier }
   minnumber 1
   maxnumber 1000
   cutoff 10.0     { distance }
   level  0.99     { fraction contribution}
   or-restraint *    { Creates new .tbl from the restraints, taking into }
                                  { account the exclusions due to violations and the }
                                  { limited assignments due to the other criteria }
  end
end


stop
```

## E-3 *check* script & *check_all_nocal.inp* script

### E-3.1 *check* script

```
setenv PROTOCOLS ../../../protocols/swapping+J
cns_solve < check_all_nocal.inp > check_all_nocal.log
```

### E-3.2 *check_all_nocal.inp* script

```
! Script to analyse ensemble of structures and the ambiguous NOE used
! to generate them in order to produce new restraints for the next
! iteration of Michael Nilges' ARIA approach
! received from Andrew Raine March, 1997

!read parameter file
parameter
  reset
  @CNS_TOPPAR:protein-allhdg.param
end

!read psf
struct
  reset
  @@../../../mtf/M2.mtf
end

!------------------------------------------------------------------------------
! Set up equivalences for methyls and aromatics etc.

  set echo off message off end
    inline @define_methyls.cns
  set echo on message on end

! To get around a bug in the noe analyse statements, we need to
! look at all the restraint files individually (actually, we could do them
! all at the same time if we were prepared to put them all into one class)


!------------------------------------------------------------------------------
! Read experimental restraints

noe
  reset
  nrestraints = 50000              ! allocate space for NOEs
  ceiling 100
end

noe
!   reset
!   nrestraints = 50000 { allocate space for NOEs }
!   ceiling    = 1000 { maximum force constant that is allowed }

   set message off echo off end
   class      c @@../restraints/filt_c-noesy.tbl
              @@../restraints/filt_NOESY_HSQC.tbl

   set message on echo on end
   averaging   c sum
   monomers    c 2
   potential   c soft
   scale       c 1 { initial scale }
   sqoffset    c 0.0
   sqconstant  c 1.0
   sqexponent  c 2
   soexponent  c 1
   asymptote   c 2.0
   rswitch     c 1.0


   set echo on message on end

   averaging  * sum
   potential  * soft
   scale      * 1.0
   sqconstant * 1.0
   sqexponent * 2
   soexponent * 1
   rswitch    * 1.0
   sqoffset   * 0.0
   asymptote  * 2.0
   ?
end
```

```
  aria
    analyze
      reset
      mode intra
    end

    countviol reset end
  end

  for $file in (
                  @files.cns
          ) loop main


  !evaluate ($count = 0)
  !evaluate ($end_count=30)
  !
  !while ($count < $end_count)  loop main
  !     evaluate ($count=$count+1)
  !evaluate ($file="refine__"+encode($count)+".pdb")

     coord init end
     coord @@$file

     aria
       analyse   { Accumulate info about all the restraints over }
                     { all the structures.  Each restraint will be associated }
                  !minimum      { with either the minimum }
                  average      { or mean }
                  !accumulate   { or r6sum value over the ensemble }
                end
     end

     aria
       countviol
                  threshold 0.5
                end
     end

     !print threshold 0.5 noe

  end loop main

  {
  {*************************************************************************}
  {This section for an analysis of each violated restraint (in the log file)}
  aria countviol
    list * 0.0 {fraction of ensemble in which restraints are violated}
                              { 0.0 means list all restraints violated at all, 1.0 means}
                              { all restraints violated in ALL structures }

  end end
  }

  {
  {*************************************************************************}
  {This section for table of nuclei contributing to each restraint}
  set print ../contributions/contrib_noexcl_all_nocal.txt end

  aria analyse          { report on the stats gathered earlier }
     minnumber 1
     maxnumber 1000
     cutoff 10.0       { distance - nothing further than this}
     level  1.0        { fraction contribution - list all possibilities making}
                 {               up to this fraction of the intensity}
     sort *          { Lists all possibilities for each restraint which     }
                 { satisfy the criteria above, in order of contribution }
                                       { to total intensity (ISPA!) }
  end end
  }

  {*************************************************************************}
  {This section for new restraint table excluding violated restraints and nuclei}
  {contributing less than a set fraction of the intensity of the crosspeak}
  set print ../restraints/check_all_nocal.tbl end

  aria
    countviol
      exclude * 1.0 { fraction of ensemble in which restraints are violated}
              { 0.0 means exclude ANY violations                }
              { 0.01 < x < 0.99 means exclude restraints violated in }
              {               more than this fraction of ensemble }
    end

    analyse          { report on the stats gathered earlier }
```

```
    minnumber 1
    maxnumber 1000
    cutoff 10.0      { distance }
    level  1.0      { fraction contribution}
    check *
   or-restraint *     { Creates new .tbl from the restraints, taking into }
                                    { account the exclusions due to violations and the }
                                    { limited assignments due to the other criteria }
  end
end


stop
```

# APPENDIX F: scripts for restraints formatting

## F-1 *fillShifts* procedure

```
#!/bin/tcsh -f

row resid le 159 < shifts_summary.rdb |\
 compute resid = resid + 1000 |\
 column resid -c resname resname.maj -c atname atname.maj \
        -c shift shift.maj -c delta delta.maj \
 >! major_summary_renum.rdb

row resid le 159 < shifts_summary.rdb >! shifts_summary_filled.rdb

row resid ge 1092 < shifts_summary.rdb >! minor_summary.rdb

nawk -f substitute1in2.awk minor_summary.rdb major_summary_renum.rdb \
 >! minor_summary_filled.rdb

headchg -del < minor_summary_filled.rdb >> shifts_summary_filled.rdb
```

## F-2 *stripAssignment* procedure

```
foreach file (NOESY_HSQC c-noesy)
 compute atname1 = null \; atname2 = null \; atname3 = null < {$file}.rdb \
 > {$file}.stripped.rdb
end
```

## F-3 *rdb2orrest.awk* procedure

```
# convert match table from connect to OR format CNS/ARIA restraints with
# inter-monomer possibilities removed
#
# usage:  nawk -f rdb2orrest.awk match_c-noesy.rdb >! or_c-noesy.tbl
#

BEGIN { xpk=-1;
}
{
 if (NR > 2) {
 if ( sqrt(($1 - $4)^2) < 70 ) {
  if ($8 != xpk) {
    xpk = $8;
    printf"ASSI {%s}\n",$8;
    printf"  ((resid %s and name %s))\n  ((resid %s and name %s))\n",$1,$3,$4,$6;
    printf"   %s  %s  0.0 peak  %s weight  0.10000E+01 volume %s ppm1 0.0 ppm2 0.0\n",$9,$9,$8,$7;
  }
  else {
    printf"OR {%s}\n",$8;
    printf"  ((resid %s and name %s))\n  ((resid %s and name %s))\n",$1,$3,$4,$6;
  }
 }
 }
}
```

## APPENDIX G: $GABA_B$ R1a CP2 list of constraints